

A COMPARISON OF REACTIONS TO PERSONALITY
AND ACHIEVEMENT TEST ITEMS

By

JOHN K. OSBORN

*Professor of Education
Hillsdale College*

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
in the University of Michigan

1486

A COMPARISON OF REACTIONS TO PERSONALITY AND ACHIEVEMENT TEST ITEMS

By

JOHN K. OSBORN

*Professor of Education
Hillsdale College*

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy
in the University of Michigan

EDWARDS BROTHERS, INC.

ANN ARBOR, MICHIGAN

1936

Copyright 1936
John K. Osborn



PRINTED IN U.S.A.

*Lithoprinted by Edwards Brothers, Inc., Lithoprinters and Publishers
Ann Arbor, Michigan, 1936*

FOREWORD

The dissertation presented in the following pages was completed in April of 1934, and was submitted at that time in manuscript form. In preparing the manuscript for publication, only such changes were made in the original format as were necessitated by the lithoprinting process. There has been no important revision of content, other than minor changes in phraseology, nor any attempt to bring the bibliographical and reference material up to date.

The writer wishes to make grateful acknowledgment to all who assisted him in any way toward bringing the project to a successful conclusion, particularly the members of the dissertation committee and other faculty members who lent their cooperation in various ways, to the students who gave generously of their time as subjects for the experimental portion of the investigation, and to the writer's wife who helped in numerous capacities.

Acknowledgment is also made to the various publishers for their kind permission to reprint the quotations appearing herein. Each quotation is also acknowledged in an appropriate footnote.

Hillsdale, Michigan
September, 1936.

TABLE OF CONTENTS

Chapter I INTRODUCTION

	Page
I. Objective Testing Techniques and Personality Tests	1
II. The Specific Problem of This Investigation	5

Chapter II THE EXPERIMENTAL PROCEDURE

I. The Group Testing	9
Subjects	9
Tests	9
Administration of the Group Tests	9
II. Selection of Items for the Individual Testing	9
III. Apparatus for the Individual Experiments	11
The Room	11
The Polygraph Recorder	11
The Psychogalvanic Response Recorder	12
The Exposure Apparatus	14
Remarks	14
Measurement of the Psychogalvanic Response	15
IV. Technique of the First Individual Experiment	16
V. Technique of the Second Individual Experiment	17
VI. Interpretation of the Records	18

Chapter III RESULTS OF THE GROUP TESTING AND OF THE FIRST INDIVIDUAL EXPERIMENT

I. Results of the Group Testing	23
Scores on the General Information Test	23
Scores on the Bernreuter Personality Inventory	23
Comparison of the Times Required for Completion of the Two Tests	24
Comparison of the Number of "?" Answers to the Items of the Two Tests	25
Comparison of the Number of Changed Answers on the Two Tests	26
Significance of the Group-Testing Results	27
II. Results of the First Individual Experiment	27
Psychogalvanic Responses in Experiment 1	27
Reaction Times in Experiment 1	30
Memory Values in Experiment 1	31
Significance of the Experiment 1 Results	31

Chapter IV PSYCHO GALVANIC RESPONSES IN EXPERIMENT 2

Subjects	32
General Results	32
Individual Subject Average Deflections	32
Individual Item Average Deflections	33
The Positional Deflection Curve, Experiment 2	33
Personality and Achievement Deflection Curves in Parts 3 and 4	35
Measures of Central Tendency	36

	Page
Reliabilities of Differences	37
Inter-type vs. Intra-type Differences	38
The Personality-Achievement Difference for the Entire Stimulus Series	38
The Personality-Achievement Difference for Parts 1 and 2 Considered as a Unit	39
The Personality-Achievement Difference for Part 3	39
The Personality-Achievement Difference for Part 4	39
The Personality-Achievement Difference for the Last Half of Part 4	39
Individual Records vs. the Group Averages	40
Relative Variabilities of Personality and Achievement Deflections	40
Average Deflections for the Individual Items	41
Effects of Sex and Vocalization of Answers	41
The Sex Difference for All Subjects	41
The Vocalization Difference for All Subjects	42
Relationships Between the Sex Difference and the Personality-Achievement Difference	43
Relationships Between the Vocalization Difference and the Personality-Achievement Difference	45
Relationships Between the Sex Difference and the Vocalization Difference	46
Mean Deflections for the Favorable, Unfavorable, "?" and Mixed Classes of Items	46
General Summary of Chapter IV	46

Chapter V

REACTION TIMES IN EXPERIMENT 2

Distribution of All Reaction Times	48
Positional Curve of Reaction Times	48
Length of Item and Reaction Time	48
Consistency of the Personality-Achievement Difference	50
Individual Subject Average Reaction Times	50
The Personality-Achievement Difference, All Subjects	50
The Sex Difference, All Reaction Times	51
The Vocalization Difference, All Reaction Times	51
Relationships Between the Sex Difference and the Personality-Achievement Difference	52
Relationships Between the Vocalization Difference and the Personality-Achievement Difference	53
Reaction Times vs. Psychogalvanic Responses	54
Reaction Times for the Four Selected Classes of Items	54
General Summary of Chapter V	54

Chapter VI

IMMEDIATE RECALL, JUDGMENTS OF "DIFFICULT" AND
CHANGED ANSWERS, IN EXPERIMENT 2

General Results	55
Effects of Serial Position on Memory for Items	55
Type, Sex, and Vocalization Differences in Recall	55
Average Recall Scores for the Selected Classes of Items	57
The Personality-Achievement Difference in Memory Scores for All Subjects	57
The Sex Difference in Total Memory Scores	58
Relationships Between the Sex Difference and the Personality-Achievement Difference	59
Relationships Between Vocalization and the Personality-Achievement Difference	60
Memory Scores vs. Psychogalvanic Responses	60
Memory Scores vs. Reaction Times	60
Significance of the Lack of Inter-relationships	60
Judgments of Difficulty	61

	Page
Changed Answers in Experiment 2	62
Summary of Chapter VI	62

Chapter VII
SUMMARY AND CONCLUSIONS

Problem	63
Procedure	63
Results	
(a) Results of the Group Testing	63
(b) Psychogalvanic Response Results	64
(c) Reaction Time Results	64
(d) Memory Value Results	65
General Summary of Results	65
Interpretation of Results	66
Problems for Further Research	67
BIBLIOGRAPHY	69
APPENDIX A. Order of Items in Belts 1, 4, 5, 2n, and 3n	71
APPENDIX B. Tables	75
APPENDIX C. Copy of the General Information Test	81

TABLES IN BODY OF THESIS

Table	Title	Page
I.	DISTRIBUTION OF SCORES OBTAINED ON THE GENERAL INFORMATION TEST	23
II.	DISTRIBUTION OF SCORES OBTAINED ON THE BERNREUTER PERSONALITY INVENTORY	23
III.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN (1) THE AVERAGE NUMBER OF MINUTES USED IN ANSWERING THE ITEMS OF THE BERNREUTER PERSONALITY INVENTORY, AND (2) THE AVERAGE NUMBER OF MINUTES USED IN ANSWERING THE ITEMS OF THE GENERAL INFORMATION TEST	24
IV.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN (1) THE AVERAGE NUMBER OF TIMES THE "?" ANSWER WAS USED IN THE BERNREUTER PERSONALITY INVENTORY, AND (2) THE AVERAGE NUMBER OF TIMES THE "?" ANSWER WAS USED IN THE GENERAL INFORMATION TEST	26
V.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN (1) THE AVERAGE NUMBER OF CHANGED ANSWERS ON THE BERNREUTER PERSONALITY INVENTORY, AND (2) THE AVERAGE NUMBER OF CHANGED ANSWERS ON THE GENERAL INFORMATION TEST	26
VI.	AVERAGE DEFLECTIONS IN EXPERIMENT 2, CALCULATED IN FOUR DIFFERENT WAYS	38
VII.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR (1) ALL PERSONALITY ITEMS, AND (2) ALL ACHIEVEMENT ITEMS	38
VIII.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR (1) THE PERSONALITY ITEMS OF PARTS 1 AND 2, AND (2) THE ACHIEVEMENT ITEMS OF PARTS 1 AND 2	39
IX.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR (1) THE PERSONALITY ITEMS OF PART 3, AND (2) THE ACHIEVEMENT ITEMS OF PART 3	39
X.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR (1) THE PERSONALITY ITEMS OF PART 4, AND (2) THE ACHIEVEMENT ITEMS OF PART 4	39
XI.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR THE LAST HALF OF PART 4: (1) PERSONALITY ITEMS, AND (2) ACHIEVEMENT ITEMS	40
XII.	INDIVIDUAL SUBJECT MEAN DEFLECTION AVERAGES FOR MEN AND WOMEN, AND FOR THE VOCAL AND NON-VOCAL GROUPS	41
XIII.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY (1) ALL MEN, AND (2) ALL WOMEN	41
XIV.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP	42
XV.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO PERSONALITY ITEMS BY (1) ALL MEN, AND (2) ALL WOMEN	43
	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO ACHIEVEMENT ITEMS BY (1) ALL MEN, AND (2) ALL WOMEN	43
XVI.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE MEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	43
	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE WOMEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	45
XVII.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO PERSONALITY ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP	45

Table	Title	Page
XVII.	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO ACHIEVEMENT ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP	45
XVIII.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	45
	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE NON-VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	45
XIX.	AVERAGE SERIAL POSITIONS AND MEAN GALVANIC DEFLECTIONS FOR THE FOUR SELECTED CLASSES OF ITEMS	46
XX.	MEAN PERSONALITY, ACHIEVEMENT AND TOTAL REACTION TIMES FOR VARIOUS CLASSIFICATIONS OF ITEMS IN EXPERIMENT 2	50
XXI.	INDIVIDUAL SUBJECT AVERAGE REACTION TIMES FOR THE MEN AND WOMEN AND FOR THE VOCAL AND NON-VOCAL GROUPS	50
XXII.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT REACTION TIMES TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	51
XXIII.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT REACTION TIMES GIVEN BY (1) THE MEN, AND (2) THE WOMEN	51
XXIV.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES FOR (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP	52
XXV.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE MEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	52
	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE WOMEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	52
XXVI.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE PERSONALITY ITEMS BY (1) THE MEN, AND (2) THE WOMEN	53
	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE ACHIEVEMENT ITEMS BY (1) THE MEN, AND (2) THE WOMEN	53
XXVII.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	53
	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE NON-VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	53
XXVIII.	(a) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE PERSONALITY ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP	53
	(b) DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE ACHIEVEMENT ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP	53
XXIX.	MEAN PERSONALITY, ACHIEVEMENT AND TOTAL REACTION TIMES FOR THE FOUR CLASSES OF ITEMS SELECTED ON THE BASIS OF THE GROUP TEST RESULTS	54
XXX.	INDIVIDUAL SUBJECT AVERAGE MEMORY SCORES CLASSIFIED TO SHOW DIFFERENCES DUE TO TYPE OF ITEM, SEX, AND VOCALIZATION OF ANSWERS	55
XXXI.	AVERAGE FREQUENCY OF RECALL OF ITEMS IN EACH OF THE FOUR SELECTED CLASSES IN EACH TYPE	57
XXXII.	DATA CONCERNED WITH THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE INDIVIDUAL SUBJECT MEMORY SCORES FOR (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	57
XXXIII.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL AVERAGE TOTAL MEMORY SCORES FOR (1) THE MEN, AND (2) THE WOMEN	58

Table	Title	Page
XXXIV.	DATA CONCERNING THE DIFFERENCE BETWEEN THE NUMBER OF PERSONALITY ITEMS RECALLED BY (1) THE MEN, AND (2) THE WOMEN	59
XXXV.	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE INDIVIDUAL MEMORY SCORES MADE BY THE WOMEN FOR (1) PERSONALITY ITEMS, AND (2) ACHIEVEMENT ITEMS	59
XXXVI. (a)	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE NUMBERS OF (1) PERSONALITY ITEMS, AND (2) ACHIEVEMENT ITEMS RECALLED BY THE VOCAL GROUP	60
(b)	DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE NUMBERS OF (1) PERSONALITY ITEMS, AND (2) ACHIEVEMENT ITEMS RECALLED BY THE NON-VOCAL GROUP	60
XXXVII.	RELIABILITY DATA CONCERNING THE DIFFERENCE BETWEEN THE AVERAGE NUMBER OF "DIFFICULT" JUDGMENTS PER SUBJECT FOR (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS	61
XXXVIII.	MEASURES ACCOMPANYING CHANGED ANSWERS, EXPERIMENT 2	62

TABLES IN APPENDIX

Table	Title	Page
I-A.	"YES", "NO", "?" AND OMITTED ANSWERS GIVEN BY 86 COLLEGE STUDENTS TO THE ITEMS IN THE BERNREUTER PERSONALITY INVENTORY	75
II-A.	"YES", "NO", "?" AND OMITTED ANSWERS GIVEN BY 86 COLLEGE STUDENTS TO THE ITEMS IN THE GENERAL INFORMATION TEST	76
III-A.	"YES", "NO", "?" AND OMITTED ANSWERS, ALSO THE JUDGMENTS OF "DIFFICULT" GIVEN BY 70 SUBJECTS TO THE SELECTED TEST ITEMS USED IN EXPERIMENTS 1 AND 2. (PERSONALITY)	77
IV-A.	"YES", "NO", "?" AND OMITTED ANSWERS, ALSO JUDGMENTS OF "DIFFICULT" GIVEN BY 70 SUBJECTS TO THE SELECTED TEST ITEMS USED IN EXPERIMENTS 1 AND 2. (ACHIEVEMENT)	78
V-A.	INDIVIDUAL SUBJECT RECORDS IN EXPERIMENT 2: MEAN PERSONALITY AND ACHIEVEMENT DEFLECTIONS, MEAN PERSONALITY AND ACHIEVEMENT REACTION TIMES, AND PERSONALITY AND ACHIEVEMENT MEMORY SCORES	78
VI-A.	INDIVIDUAL SUBJECT RECORDS IN EXPERIMENT 1: MEAN PERSONALITY AND ACHIEVEMENT DEFLECTIONS, MEAN PERSONALITY AND ACHIEVEMENT REACTION TIMES, AND PERSONALITY AND ACHIEVEMENT MEMORY SCORES	79

FIGURES

Number	Title	Page
1.	Floor Plan of the Experimental Room	12
2.	Electrical Circuits for the Polygraph Recorder	13
3.	External Battery Circuit for the Psychogalvanograph	14
4.	Sample Portion of a Record Sheet from the First Individual Experiment . .	19
5.	Sample Record Graph from the Second Experiment	20
6.	Times Required for the Completion of the Bernreuter Personality Inventory and the General Information Test	25
7.	Frequency of "?" Answers Given by 88 Subjects to the Items of the Bernreuter Personality Inventory and the General Information Test . .	26
8.	Distribution of Changed Answers in the Bernreuter Personality Inventory and the General Information Test	27
9.	Positional Deflection Curves, Experiment 1. (Belts 1, 4, and 5 Combined)	29
10.	Distribution of All Deflections, Experiment 2	32
11.	Distribution of Individual Subject Mean Deflections, Experiment 2 . . .	33
12.	Distribution of Individual Item Mean Deflections, Experiment 2	33
13.	Positional Deflection Curve, All Subjects, Experiment 2	34
14.	Mean Deflection Curves, Belts 2n and 3n, Experiment 2	35
15.	Positional, Personality, and Achievement Deflection Curves, Parts 3 and 4, Experiment 2	36
16.	Smoothed Deflection Curves (From Figure 15)	37
17.	Distribution of Individual Mean Deflections for Men and Women	42
18.	Distribution of Deflections Given by the Vocal and Non-Vocal Groups . .	42
19.	Distribution of Individual Mean Deflections for the Vocal and Non-Vocal Groups, Experiment 2	43
20.	Positional Deflection Curves for Vocal and Non-Vocal Groups Compared with Positional Deflection Curve for All Subjects, Experiment 2 . .	44
21.	Distribution of All Reaction Times Accompanying the Responses to Personality and Achievement Items, Experiment 2	48
22.	Average Reaction Time for Each Serial Position, Experiment 2	49
23.	Distribution of Individual Subject Average Reaction Times, Experiment 2	51
24.	Distribution of Individual Average Reaction Times for Men and Women, Experiment 2	51
25.	Individual Subject Reaction Times for the Vocal and Non-Vocal Groups, Experiment 2	52
26.	Average Memory Value of Each Serial Position, Experiment 2	56
27.	Distribution of Individual Subject Memory Scores, Experiment 2	58
28.	Total Memory Scores of Men and Women, Experiment 2	58
29.	Distribution of Personality Item Memory Scores for Men and Women, Experiment 2	59
30.	Distribution of Personality and Achievement Memory Scores for Women, Experiment 2	59
31.	Distribution of Judgments of "Difficult" Accompanying the Responses to Personality and Achievement Items	61

ILLUSTRATIONS

Number	Title	Page
1.	Lower deck of Table T ₃ , showing the polygraph recorder, kymograph, paper supply roll, battery and metronome box	21
2.	Close-up view of the polygraph recorder, showing the pens and the sliding pencil-point carriage	21
3.	The Wechsler Psychogalvanograph as used in this experiment	21
4.	Upper deck of Table T ₃ , showing the glass scale, sliding pointer, colored bulbs, and control instruments for operating the Psychogalvanograph	21
5.	Subject's chair: (a) with exposure apparatus in place, and (b) with exposure apparatus lifted out of the way	21
6.	Exposure apparatus, subject's lamp, answers keys, and electrodes: ensemble view	21
7.	The electrodes and platform for supporting the left hand, with hand in place	22
8.	Subject's view of the exposure apparatus and instruction sheet	22
9.	The experimenter in working-position for Experiment 2	22

Chapter I

INTRODUCTION

I. Objective Testing Techniques and Personality Tests.

The purpose of this investigation is to determine whether there are differences between the reactions given by a group of subjects to a number of "Yes-No" questions of the kind frequently found in objective or new type achievement tests and the reactions of the same group of subjects to an equal number of "Yes-No" items taken from the Bernreuter Personality Inventory, the reactions being observed under conditions such that any obtained differences may be reasonably attributed to the respective subjective effects of the two kinds of questions (i.e., achievement and personality).

The problem just stated in general terms is one aspect of a much wider issue which has been the basis for extensive controversy during recent years, namely, whether personality characteristics may be tested accurately by means of the same pencil and paper techniques which have been found to be reasonably valid and reliable for the measurement of school achievement and of abstract or verbal intelligence. The present investigation does not attempt to obtain the final answer to this wider problem, but only to explore the potentialities of one experimental approach to one of the many subordinate issues which are involved in the main problem.

The testing movement, as a distinct branch of educational theory and practice, may be regarded as having consisted to date of three principal phases, namely, achievement testing, intelligence testing, and personality or character testing. These phases overlap extensively both in content and in temporal order, yet exhibit a certain degree of evolutionary sequence.

The testers¹ concentrated much of their earlier effort on the measurement of achievement in the school subjects. Rice, whose investigation using the comparative test (1894-1895) marks a turning-point in educational thought, is usually considered to be the pioneer (in America, at least) of the

achievement testing phase of the testing movement; but the first real impetus came during the initial decade of the present century. That period saw the publication, in 1904, of Thorndike's Mental and Social Measurements, and such outstanding pioneer measuring instruments as the Thorndike handwriting scale and Series A of the Courtis arithmetic tests (both in 1909). The major contribution resulting from the endeavors to measure school achievement has been the standardized test, an examination the results of which can be interpreted in terms of norms or standard scores (averages obtained by giving the test to large numbers of individuals under comparable conditions).

Toward the close of the second decade of the century, the testers shifted the center of their attention to the measurement of intelligence by group testing methods, seeking greater economy of time and effort than was possible with the use of the individual testing methods developed by Alfred Binet. Binet's 1905 scale had made such testing a reality after various earlier laboratory attempts by Cattell and others (beginning about 1890) had proved largely futile because they dealt chiefly with reaction time, sensory discrimination, and similar functions, rather than with the more complex processes in which differences in capacity are most readily observed and measured.²

The Army Alpha test (authorized by the Surgeon General early in 1918 and published for Army use later in the same year) brought to a head and gave wide publicity to this second phase of the testing movement, the result being a score or more of widely used group tests of intelligence and several times that many of lesser fame. The National Intelligence Test (copyrighted in 1920) was one of the earliest of the better known group intelligence tests. These tests, produced mostly during the decade following the War, contributed, or at least helped greatly to popularize, a most important advance in testing technique, the objective or new type question. Army Alpha, for example, utilized such

¹The term "testers" is meant to refer not to everyone who has dealt with any sort of test, but rather to that group of psychologists and educators (mostly Americans) whose efforts have been mainly responsible for the flood of tests and scales on the market for the last two decades.

²Binet had also been working with the problem as early as 1895, but the 1905 scale was the first practical result of his efforts, and by that time the peak of interest and activity of the other sort mentioned above had been reached and passed.

item forms as multiple choice, true-false, best answer, same-opposite and analogy. The achievement testers at once sensed the possibilities of the new type item and seized upon it avidly. The educational literature of the third decade of the century was replete with opinions and experimental evidence regarding the relative merits of the new type test versus the older essay test. The result was a decisive victory for the new type examination in the realm of published standardized tests, as well as extensive popularity for the "informal" objective test, constructed by the user as needed. The inadequacy of the new type examination for testing many desirable products of learning was quite generally recognized, but on the whole, the new type test item has become firmly entrenched both in achievement testing and intelligence testing by methods which employ pencil and paper techniques.

The use of similar testing techniques in both fields was natural, since the group intelligence examination, to a considerable extent, actually tested achievement and only inferred intelligence from the scores obtained on what were, largely, achievement tests. Colvin (5)³ stated this point clearly:

An inspection of the various intelligence tests now in use clearly shows that psychologists have accepted this definition [intelligence as ability to learn] practically, if not theoretically. Intelligence tests are by no means confined to problem-solving, even in its simplest forms. They determine an individual's intelligence largely in terms of what he has learned, thus obtaining a measure of his ability to continue learning. Vocabulary tests, range of information tests, same-and-opposites tests, tests of fundamental operations in arithmetic (one of the most widely used) and the like, demand very little that is novel, little that tests rational powers. If an individual has sufficient knowledge and skill he can pass these tests. They measure intelligence only on the assumption that they test ability to learn by discovering what has already been learned. Even those tests that involve ingenuity, deliberation, and choice with words or things are based on elements that show what a person has already acquired.⁴

The assumption was also made, tacitly if not openly, that the individuals being tested would have had substantially equivalent opportunities to achieve the knowledges

and skills measured by the intelligence test, and that the levels of achievement actually reached would therefore be proportional to the intellectual capacities of the individuals reaching those levels.

The intelligence test became the center of an active controversy concerning its merits, the reverberations of which, while considerably abated, are still being heard. The testers meanwhile moved on to greener pastures and ushered in the third phase, that of personality or character testing. Manson (14) listed 1364 references in this field up to 1926, and the reviews which have appeared almost yearly since 1926 in the Psychological Bulletin have swelled the total impressively. The 1933 review by Watson (23) listed over 200 titles which had appeared since the compilation of the immediately preceding review (1932), even after excluding rating scales, physiological measures, and one or two other lines of endeavor usually included in the general field.

If one accepted uncritically any recent list⁵ of personality tests now available in published form, he might easily conclude that it is possible to measure almost any human characteristic. An increasingly large number of writers, however, hold the opinion that the drawing of such a conclusion from this abundance of tests is, to say the least, premature. The following quotation from Gilliland (9) is typical of the critical comments which have been made:

The construction of tests of personality is the most popular psychological pastime of the last half-dozen years....These tests include measurements of everything from facial expression to morality and religion. If one in a dozen were worth the time and effort required to give it we could have as full a reproduction of personality in quantitative terms as we now have of a modern real estate development. Unfortunately...success is not in direct proportion to the number of tests in the field.⁶

The interest of the present investigation in the historical sequence which has been traced above arises from the fact that many of the personality tests which have been constructed represent attempts to carry over into this newer field the same techniques used in measuring achievement and intelligence; that is, the tests are objective pencil and paper tests, making extensive use of new type test

³Numbers in parentheses following an author's name or other indication of a direct quotation or reference refer to the items in the numbered bibliography following the body of the thesis.

⁴Colvin, S. S. "Principles Underlying the Construction and Use of Intelligence Tests", Intelligence Tests and Their Use, Part I, Twenty-first Yearbook of the National Society for the Study of Education, 1923. Pp. 16-17. Quoted by permission of the Society.

⁵Such as Hildreth's bibliography (11), or the Review of Educational Research number devoted to personality and character testing (5), or the list in the Tenth Yearbook (16).

⁶Gilliland, A. R. "Problems of Personality." Journal of Abnormal and Social Psychology, XXIII (Oct.-Dec., 1928). Pp. 373-4. Quoted by permission of the publishers, the American Psychological Association.

items. Raubenheimer (17) noted and approved this tendency quite early:

A great many of the practices that have been developed and standardized in the field of intelligence testing and educational measurement are of service in connection with the development of character-behavior tests....Many of the test forms—performance, completion, underscoring, etc.—and administration devices may be utilized. And lastly, many of the scoring and statistical methods may lend themselves to ready use in this other field.⁷

The foregoing statements regarding the borrowing of techniques from the achievement testing field apply especially to those tests and questionnaires of the self-description sort, in which the testee is asked directly for information about his habits, likes and dislikes, fears, attitudes, etc. Inspection of the ordinary run of personality tests of this type which have been published in pencil and paper forms shows that many of the standard objective item forms have been used extensively. True-false and multiple choice items, in all of their many variant forms, are the kinds most frequently employed, apparently on the assumption that the possession of information about the self may be tested in the same way as the possession of information about any factual subject-matter.

It is probably true that the best way to obtain data regarding the characteristic behavior traits of an individual is through direct observation of that individual in a large variety of situations where the traits to be observed can operate free from distorting influences. As Symonds (21) points out, however, there is a natural temptation to short-cut these tedious observational methods by asking the individual about himself. Symonds continues by discussing some of the well-recognized dangers in connection with this procedure:

This substitution of testimony for direct observation carries with it certain implications....To ask a child about his interests, [etc.]....is to ask him about things that he perhaps has never observed. The greater part of a person's conduct originates without the aid of his verbal organization....Along with the question of the possibility of self-observation goes the matter of the voluntary alteration of language to fit purposes....There may be only the loosest connection between language and conduct, between what he says he believes or likes and what he actually believes or likes as shown by his conduct.⁸

And again, a few pages earlier:

In the questionnaire, however, where the pressure of time is removed, there is a possibility that answers may be given for the effect they produce on the examiner. If the questionnaire permits the individual answering it to show himself to advantage, then there will be every incentive to tell the truth. But if truthful answers reveal things which may be condemned or criticized, then the temptation to falsehood becomes very strong.⁹

Many criticisms of this general sort may be found in the literature. A few samples follow. Thomas and Thomas (22) in their work on the study and treatment of the problem child said:

There has been, during the past ten years, a flood of literature describing the various attempts to measure personality differences, to define the various components of character, and to measure the relationships between these various factors and behavior. The difficulties facing those who attempted to measure these factors have been even greater than those facing the mental testers.

The primary difficulty has been to make such definitions of the traits to be measured as to give any degree of reliability to the measurements....Furthermore, no general agreement (even general agreement among the workers in the field) can be obtained as to the test situations which will be valid indications of the trait in question. Great difficulties have further arisen with regard to the generality of any trait as determined by any test situation or paper and pencil test.¹⁰

The following comment of Allport and Vernon (2) is pertinent to the general problem giving rise to this investigation:

...there seems to be an increasing dissatisfaction with the more artificial pencil and paper or laboratory methods, together with the realization that, while the test situation should be controlled, yet both situation and response should be as natural as possible....Apparently the more strictly objective the approach, the less normal and spontaneous are the testee's reactions, and therefore the less significant the results. There has been too great a tendency to ascribe such poor results to lack of consistency and integration in the testees rather than to defects in the techniques.¹¹

A recent yearbook on character education (16) also criticizes those who would test character merely by extending existing techniques:

Character measurement has some of the advantages and problems of a younger child in a family which

⁷Raubenheimer, A. S. "An Experimental Study of Some Behavior Traits of the Potentially Delinquent Boy", *Psychological Monographs*, XXXIV (1925, No. 5). P. 8. Quoted by permission of the publishers, the American Psychological Association.

⁸Symonds, P. M. *Diagnosing Personality and Conduct*. Pp. 17-18. Quoted by permission of the publishers, the D. Appleton-Century Co.

⁹*Ibid.*, p. 14.

¹⁰Thomas, W. I., and Thomas, D. S. *The Child In America*. Pp. 370-71. Quoted by permission of the publishers, Alfred A. Knopf, Inc.

¹¹Allport, G. W., and Vernon, P. E. "The Field of Personality", *Psychological Bulletin*, XXVII (Dec., 1930). P. 703. Quoted by permission of the publishers, the American Psychological Association.

has made a name for itself. On the one hand, tests in other fields have demonstrated the values in measurement....On the other hand, difficulty is apt to arise when expectations for a younger child are set by the accomplishment of older siblings. Character tests cannot be expected to fit precisely the patterns set by other forms of measurement. They will have their own limitations and their own possibilities....The measurement [i.e., of an attitude] may still be important, but its nature and interpretation will not correspond exactly to achievement tests in subject-matter fields.¹²

Through all of these criticisms, and others which might be cited, there runs a strong note of skepticism regarding attempts to test personality in the same way that achievement and intelligence have been tested in the past. This skepticism has found concrete expression in various attempts to formulate other measuring techniques, such as rating, direct observation, use of physiological measures, etc. Most of the latter methods, however, have the same disadvantages as the individual intelligence test, that is, they are too laborious and time-consuming. Hence the question of the applicability of objective pencil and paper methods to the measurement of personality traits is by no means a dead issue.

It is to be emphasized that the worst of the difficulties touched upon in the preceding quotations are largely subjective in nature, especially as they concern the self-description tests and questionnaires. The yearbook (16) already quoted recognizes this fact:

Any individual's description of his own character, his symptoms, and his behavior under certain imagined circumstances is of value if: (1) he knows the truth about himself in this matter, and (2) he cares to tell it, and (3) the instrument records and interprets it without ambiguity.¹³

These three conditions of satisfactory test performance may easily be generalized and applied outside of the personality testing field. As far as the achievement test is concerned, the conditions are reasonably well met. The testee generally has had more or less definite opportunity to learn the facts or skills in which he is being tested. Ordinarily he does desire to tell what he knows, and to have what he tells correspond with the actual truth as closely as possible. Likewise, if it has been well constructed, the test may be relied on as an instrument really capable of revealing the testee's true status with re-

gard to the skill or information being tested.

While all three of these conditions are important, the concern of this study is chiefly with the second, or motivational factor. "Making a good showing" in the case of the achievement test clearly necessitates conformity with the prescribed directions, as well as strict adherence (within the limits of the testee's knowledge) to the real truths of the thing being measured.

In the case of the personality test, however, the subjective dynamic factors in the testing situation are easily reversed. Confronted with the question, "Did Caesar cross the Rubicon?", the testee will normally do his best to recall his real knowledge of the situation, and will answer according to what he knows or believes to be the truth. On the contrary, confronted with the question, "Are you bashful?", the testee may be influenced by a whole host of complexes, fears, suspicions, feelings, emotions, rationalizations, etc., and the answer given may or may not be the truth as the testee sees it. Yet the only external difference between the two questions lies in the personal reference of the second. The question, "Is Al Capone a criminal?", may be answered very objectively and impersonally. The question, "Do you ever steal?", is quite another matter.

The question of honesty, therefore, is of primary importance. The testee may know the desired personal fact very clearly, and the test item may be perfectly phrased to reveal the fact; but if the subjective dynamic factors are not favorable, the net result is liable to be a falsehood or an evasion. Honesty is certainly influenced by the number and intensity of the subjective conflicts which the act of responding to a test item arouses in the testee.

Many critics of the adjustment questionnaire and other self-describing personality tests feel that the omnipresent possibility of dishonest answers is sufficient reason for rejecting such tests entirely. Symonds (21), however, takes what seems to be a much more reasonable position:

Instead of condemning questionnaires in a wholesale fashion because they may yield dishonest evidence, we should assume the attitude that here is a disturbing element which must be studied and its influence discovered under varying conditions so that it may be controlled or allowed for.¹⁴

Clearly, the whole situation with

¹²Character Education. Tenth Yearbook, Department of Superintendence, National Education Association. P. 345. Quoted by permission of the Department.

¹³Ibid., p. 362.

¹⁴Symonds, op. cit., p. 19.

regard to the validity of the pencil and paper personality test offers a field for experimental exploration and for the collection of definite evidence to confirm or deny the implications of the theoretical criticisms. Among the first points of attack, it would be desirable to demonstrate whether or not what have been called the subjective dynamic factors actually do operate differently in the case of the personality test item than in the case of the achievement test item. It is essentially the task of this investigation to attempt to secure evidence relating to one aspect of this phase of the wider problem.

As thus stated, the task is clearly exploratory in nature, rather than a specific determination of relationships between sharply defined alternatives. There must first be sought an answer to the "whether", an answer which may be in part negative, before it will be possible to indulge in any detailed consideration of the "what." It is to be hoped, however, that this study may contribute in some degree to the latter as well as to the former.

In summary: the transfer of objective techniques from the fields of achievement and intelligence testing to the field of personality measurement has raised a storm of criticism, in part because of the chances for dishonesty and other disturbing reactions believed to be present in the personality testing situation to a greater degree than they are present in the achievement testing situation. Whether or not these criticisms invalidate the use of pencil and paper techniques in personality measurement is a problem for research rather than merely for discussion. The present investigation proposes, in general, to attack one limited phase of the problem, namely, whether the subjective conditions surrounding personality testing differ from those surrounding achievement testing.

II. The Specific Problem of This Investigation.

The problem stated somewhat broadly in the preceding section of this chapter must now be restated in terms of the specific phenomena to be studied and the techniques for studying them.

Repeated mention has been made of the subjective dynamic factors in the testing situation, a term which calls for further explanation. The adjective dynamic is used here in its more or less accepted meaning in modern psychology, that is, as referring to the motivational or causal aspects of behavior. In the older terminology, dynamic psychology might be said to be concerned with the affective and conative aspects of active behavior,

rather than with the relatively static cognitive processes which were almost the sole concern of the previous generation of introspective psychologists.

Subjective dynamic factors in the testing situation are any and all states of mind in the testee which influence his response to a test item, aside from a purely impersonal cognition of what that response ought to be. Fears, complexes, emotions, general or specific reaction tendencies, wishes, inhibitions, etc., are all included in the term. In general, the favorable subjective dynamic factors are considered to be those characterized by the absence of disturbances or conflicting motives which would tend to obscure or distort the influences of the purely cognitive factors. Unfavorable subjective dynamic factors are those which are productive of sufficient conflict to interfere with the direct manifestation of the testee's impersonal knowledge concerning the facts of the matter touched upon by the test item.

If ideal testing conditions could be established, the testee's responses would be determined only by purely impersonal considerations, regardless of the nature of the test items. The common belief that such objectivity is more closely approximated in the response to the achievement test item than in the response to the personality test item has been noted. Thus, to recall a previously used example, the item, "Did Caesar cross the Rubicon?", is, for the average testee, simply a matter of impersonal historical fact. His efforts to answer the question will ordinarily involve only an honest review of what he is able to recall regarding that particular portion of Roman history, and the answer finally given will usually represent his best judgment as to the actual truth of the situation. The judgment may not be correct, but at least it will not be a conscious lie. The testee's own self-estimation and scale of personal values seldom enter into the determination of the answer to be given to such an item. The testee is not making out a case for himself, directly, and while he may know that unpleasant consequences result from a wrong answer, he also knows that those consequences cannot be avoided through intentional falsification. On the other hand, exactly the reverse of the above described situation is likely to occur when the item has a personal reference and an unavoidable connotation of right or wrong behavior or standards.

Impersonal objectivity, of course, is not entirely improbable in the case of such an item as, "Are you afraid of the dark?". The testee might, and often undoubtedly does,

review his experiences quite dispassionately and answer according to what he knows or believes to be the truth of the matter. As in the case of the achievement item, the answer may not be true, but at least it would not be a deliberate falsification of the sort implied by most of the criticisms levelled against the personality test.¹⁵

According to the theoretical argument against the personality test, however, the chances are that less favorable factors will enter into the determination of the answers to many of the questions like the one above. The testee may be acutely afraid of the dark, and realize that fact, yet also realize that he should not be, and feel unwilling to reveal that he is. Thus there arises a conflict between truthfulness and self-interest, and the response (or so say the critics of the personality test) is probably influenced fully as much by the outcome of this conflict as by what the subject knows or believes to be the actual facts in the case.

The problem may now be partially restated as follows: are the responses to personality test items, on the whole, characterized by the presence of unfavorable subjective dynamic factors to a greater degree than are the responses to achievement test items? The general consensus of opinion in the quotations cited in the preceding section of this chapter would seem to favor an affirmative answer to this question. It is the purpose of this investigation to attempt to obtain objective evidence bearing upon certain aspects of this particular implication of the theoretical criticisms of the paper and pencil personality test. Lest the accusation of "elaboration of the obvious" (always an easy charge to hurl at the investigator who seeks to test established opinion for its soundness) be raised, let the conviction be stated that the fact that many may consider this matter already settled satisfactorily on a priori grounds does not make it any the less a valid problem for research.

The above partial restatement of the problem is still rather general, and must be further simplified, this time in terms of technique. The immediate task is double. There must first be selected measuring instruments to use in securing indices of subjective processes, and then there must be set up an experimental situation in which the effects of

responses to the two kinds of items on these subjective processes can be compared directly with other variables controlled or allowed for. The first of these tasks properly precedes the further and final statement of the specific problem; the second will be considered in detail in the chapter on procedure.

The technique that has been adopted for comparing the effects of responses to personality and achievement test items is original only in the sense that apparently it has never been applied to this specific problem. Briefly stated, it is essentially a variation of the association-reaction, psychogalvanic response, memory-value combination used by Smith (20) in his studies of emotion. So far as this study is concerned, the importance of Smith's work is that it demonstrated that these measuring instruments may be used to detect differences in the subjective processes accompanying the responses to association-word stimuli. Whether the differences are primarily emotional, as Smith believed, and whether or not his much disputed theory of the bidirectional relationship of affectivity and memory is to be accepted, are questions which are not germane to the present investigation.

Studies by Jones and Lynch (12, 13) offer confirmation of those aspects of Smith's findings which are pertinent to this investigation, although the two investigators just named reached opposite conclusions regarding the bidirectional theory. Jones used 16 association words selected from Smith's list, and his results, both for psychogalvanic responses and for memory values agreed substantially with Smith's results. Jones' study is especially pertinent, inasmuch as the memory values were determined by immediate reproduction, the method which it is here proposed to use. In Jones' study, as in Smith's, the words fell into "critical" and "non-critical" groups, with distinctive psychogalvanic responses and memory values.

Lynch dealt only with the problem of memory values, although he related his results to the psychogalvanic responses obtained by Smith and Jones for the same words, and found fair correlations. Lynch tested the memory values by the recognition method, obtaining results agreeing well with Jones' results.

Such studies as the foregoing are based on the assumption that the subjective

¹⁵It should be noted that the attempt is here made to distinguish sharply between the untrue answer and the dishonest answer. If the testee really believes a certain thing to be true, even though it is actually false, and answers according to that belief, then his answer is not dishonest in the sense meant by the critics of the personality test. Therefore the subjective situation would be different, supposedly, from that accompanying a conscious falsification, where the subject believes one thing, but answers as though he really believed the reverse.

conditions underlying responses to association words can be differentiated in terms of psychogalvanic deflections and memory values. Reaction times, especially unusually long or short ones, have long been considered as complex indicators in this type of experiment, although the evidence is conflicting concerning the relationship between psychogalvanic deflections and reaction time. Hathaway (10) claimed to have found a correlation of .60 between the two measures. Crosland and Beck (6), more recently, deny Hathaway's claims. At all events, the point is not vital as far as this investigation is concerned, and perhaps it may be dismissed with Symonds' statement (21):

The psychogalvanic reflex promises to be of service in interpreting the reaction experiment by indicating whether failure to respond or a long reaction time is due in the particular case more precisely to intellectual difficulty in interpreting the word or to a conflict of two or more innocuous stimuli, or is really accompanied by emotion. It might also be serviceable in the case of a prolonged reaction time to indicate whether or not implicit reactions with an emotional coloring are aroused, though they pass without overt expression.¹⁶

Two other studies may be mentioned briefly in connection with this matter of detecting subjective differences, both studies relating to psychogalvanic responses. Darrow (7) found that various sorts of ideational stimuli, including brief, phraselike questions, could be classified as "disturbing" or "indifferent" on the basis of psychogalvanic deflections. Abel (1) found:

...a fairly definite functional connection between electrical changes in the skin and certain reportable attitudes (Bewusstseinslagen) which mark sudden, decided and momentary checks in the course of the comprehension and solution of simple problems....¹⁷

If the assumption may now be made (as seems to be reasonable) that responses to ideational stimuli in general are probably influenced by subjective factors similar to those which influence the association-reaction responses, then the applicability of the combined psychogalvanic response, reaction time, memory value technique to this problem is direct and obvious. The task is then to determine whether or not the effects of the responses to the two kinds of test items, measured with regard to the above three agents, differ significantly with respect to each

other, or perhaps by amounts which are greater than differences between classes of items of the same kind. If such differences are found, then it may be assumed that the subjective conditions influencing the responses also differ, and conversely.

The problem may now be stated definitely and specifically: Are there any differences between the reactions to personality test items and achievement test items, viewed in the light of their respective effects on psychogalvanic responses, reaction times, and memory values?¹⁸

Obviously, a complete answer even to this question (much less the larger issues from which it has been taken) could be made only after many researches of the general sort here proposed. There are many different kinds of achievement test questions and also many different kinds of personality test questions. Likewise there are many different sorts of individuals who might be used as subjects, and many different conditions under which the necessary testing might be done. A large number of different combinations of these variables would have to be tried before any final answer might be given. This study will attempt only a beginning, with one of the many possible combinations.

While the attempt has been made to delimit the problem rather narrowly in the foregoing pages, perhaps an extra word of caution regarding the actual scope of the present study would not come amiss at this point. Much of the discussion has been concerned necessarily with the larger and more remote bearings of this investigation rather than with its expected immediate outcomes. The study is not expected to offer any answer to the main question regarding the general validity of paper and pencil tests of personality traits, the discussion in this connection having served simply as a point of departure in orienting the specific problem. Nor will it be possible to make any sweeping generalizations about personality test questions and achievement test questions as whole classes, the findings being limited to a few examples of each kind. Likewise, there will be attempted very little qualitative analysis of what have been called the subjective dynamic factors in the testing situation. The study will be concerned directly only with the question as to whether the responses to certain selected items from

¹⁶Symonds, op. cit., p. 431.

¹⁷Abel, T. M. "Attitudes and the Galvanic Skin Reflex", *Journal of Experimental Psychology*, XIII (Feb. 1930), p. 60. Quoted by permission of the publishers, the American Psychological Association.

¹⁸Certain other comparisons not included in this statement will also be made, but the major interest will be in the three measures here named.

one personality test and one achievement test affect differently certain measures commonly believed to be influenced by subjective conditions. If significant differences are found (and it would be too much to expect positive results at all points touched), there can be only speculation as to the ultimate meaning in terms of the larger issues from which this limited problem springs.

The main point to be kept in mind is that the approach to the larger issues giving rise to this study should be experimental rather than merely rhetorical, and should be based on demonstrated fact rather than on professed belief, however authoritative the latter may be. A small contribution to the mountain of fact necessary for deciding the ultimate issues is all that is attempted in the present study.

Finally: if personality tests of the paper and pencil variety are guilty of the high crimes charged against them, one major cause surely must lie in the different subjective forces at work in the personality testing situation as compared with the achievement testing situation. These forces are presumably more disturbing in the former case than in the latter, because of the personal reference which is involved. If, however, they are large enough to justify the criticisms levelled against the tests, their effects ought to show up in connection with such measures as those which have been chosen for use here. Therefore it is worth while to try to find out whether selected items from the two kinds of tests will show the differences that the theory leads us to suspect. The latter matter alone is the immediate concern of this investigation.

Chapter II

THE EXPERIMENTAL PROCEDURE

I. The Group Testing.

Subjects. The subjects used in this investigation were 88 students enrolled in the elementary course in Educational Psychology, or in the Correlated Course¹ in Education, at the University of Michigan School of Education for the first semester of the academic year 1933-34. There were 27 men and 61 women. Ages ranged from 18 to 40. The median age was 22, with 58 cases falling in the 20-24 interval, and only 3 cases above age 29. The subjects were Juniors and Seniors, with a small sprinkling of Graduate students. All were candidates for teachers' certificates.

Tests. For the purposes of this investigation, the Bernreuter Personality Inventory was chosen as a representative of one fairly large and prominent class of pencil and paper personality tests. The Inventory is one of the more recent and more carefully constructed descendants of the Woodworth Personal Data Sheet, and has the advantage of yielding four different measures: neurotic tendency, self-sufficiency, introversion-extroversion, and dominance-submission.² It consists of 125 items adapted largely from earlier questionnaires. Each item may be answered "Yes", "No" or "?". The answers are weighted separately for each item with reference to their relative importance as indicators of each of the four traits. The score, computed separately for each trait by means of specially devised scoring stencils, is the algebraic sum of the weights. Percentile norms are furnished for high school, college and adult groups, with separate norms for the two sexes in each group.

As representative of the kind of questions found in one kind of achievement test, a General Information Test was specially constructed for use in this investigation. In general form it paralleled the Bernreuter test, consisting of 125 items, each of which could be answered by encircling "Y", "N" or "?". The true-false item does not ordinarily offer a "?" alternative, but it had to be included in this case in order to make the re-

sponses comparable with those given on the Bernreuter test. The items were selected from several different subject-matter fields, including literature, history, science, mathematics and current events, such a test being deemed more suitable for this study than a test limited to any one field of knowledge. Many of the items were adapted from published tests in various subjects. Those items dealing with psychology and current events were devised by the writer, and a few miscellaneous items were based upon facts taken from the Questions and Answers department of a popular magazine. A copy of the test will be found in the Appendix.

Administration of the Group Tests.

Through the courtesy of the instructors, the duty of acting as subjects in this investigation was included among the required laboratory exercises given in connection with the work in elementary Educational Psychology. Both group tests described above were given during a single laboratory period. The writer gave all of the tests personally, and endeavored to hold conditions as constant as possible in the various classes (5 in all) which were tested. The group testing was done during the second week of the semester, in order to allow ample time for analysis of the scores before beginning the individual tests. In each class, brief remarks relative to the general purpose of the testing were first made. The Bernreuter test was then given, following the instructions in the accompanying manual. The General Information Test followed. No time limits were employed for either test, but the subject was required to record the time of beginning and finishing the test, in order to furnish a measure of the speed of work. Announcement was also made at this time concerning the individual test which would come later, and each subject was asked to indicate on one of the blanks the hours which he had available for this purpose.

II. Selection of Items for the Individual Testing.

¹The Correlated Course requires the student's entire time for a semester, and covers the subject-matter of all of the professional courses leading to the teachers' certificate.

²Since this study was made, the Inventory has been extended by the addition of scales for confidence in oneself and sociability.

The responses to each item of each test were tabulated separately under the headings "Yes", "No", "?" and "Omitted." (Appendix, Tables I-A and II-A). The items to be used in the individual testing were selected on the basis of these tabulations. Partly because of the time involved, and partly because of certain technical limitations applying to the individual testing, it was necessary to use a relatively short list of items in this second part of the investigation. It also seemed desirable to employ items which had already been differentiated into fairly distinct classes, in order that it might be possible to some extent at least to study differences within the types as well as differences between the types.

One class was the same for both types, namely, those items which received the largest percentages of "?" answers. This basis of classification was not very satisfactory in the case of the personality items since, as will be discussed in more detail later, the "?" answer was used very infrequently on the Bernreuter blanks. However, the five items were chosen from each test which met the stated requirements.

Another scheme of classification was based on whether the answer affected the total score favorably or unfavorably. Applied to the General Information Test, this standard called for the selection of the five items answered correctly by the highest percentages of the subjects, and also the five items answered incorrectly by the highest percentages of the subjects. These items were called respectively the Easy and Difficult classes. In the case of the Bernreuter test, the items were scored according to the BI-N scale (neurotic tendency), and the choices consisted of the five items answered "normally" and likewise the five items answered "neurotically" by the largest percentages of the subjects. These items are called respectively the Normal and Neurotic classes.

A fourth class of five items from each test was chosen from those items whose answers showed no leanings toward favorable, unfavorable, or "?" answers. These classes are called Mixed.

The total number of items was 40, 20 from each test. They are listed below. The numbers before the items are their numbers on the original group test blanks.

Easy

9. Was Dickens interested in the conditions of the poor?
17. Is the base of a triangle always the longest side?

59. Are proteins a necessary food element?
110. Was "The Gold Bug" a tale of the California gold rush?
122. Does a pastoral usually deal with country or farm life?

Difficult

1. Is Valley Forge situated in New Jersey near Trenton?
29. Can tobacco be cultivated successfully in Northern States?
95. Was St. Paul one of the original twelve Disciples?
101. Is the word "supercede" spelled correctly?
109. Does USSR mean United Soviet States of Russia?

Mixed (Achievement)

32. Was President Garfield shot by the assassin Czolgosz?
48. Are spiders classified in the same phylum with crabs?
65. Does union labor favor the "closed shop" policy?
82. Is Bogota the capital of the Republic of Venezuela?
112. Are postulates proved in geometrical reasoning?

"?" (Achievement)

34. Is the anteater an edentate animal?
35. Is Popocatepetl in the same country as Chichen-Itza?
36. Are there any nudists among the Dukhobors?
54. Is Wolfgang Köhler one of the better known psychoanalysts?
94. Is the mesencephalon a neurological structure?

Normal³

3. Do you usually work things out for yourself rather than get someone to show you?
10. Are you easily discouraged when the opinions of others differ from your own?
29. Do you ever heckle or question a public speaker?
67. Do people ever come to you for advice?
81. Does admiration gratify you more than achievement?

Neurotic

2. Do you daydream frequently?
4. Have you ever crossed the street to avoid

³Items from the Personality Inventory are quoted by permission of the publishers, the Stanford University Press.

meeting somebody?

37. Do you usually try to avoid bossy or dictatorial people?
89. Do you like to get many views from others before making an important decision?
100. Do you prefer to be alone at times of emotional stress?

Mixed (Personality)

11. Do you try to get your own way even if you have to fight for it?
30. Do you very much mind taking back articles you have purchased at stores?
78. When you are in low spirits do you try to find someone to cheer you up?
88. If you came late to a meeting would you rather stand than take a front seat?
90. Do you try to treat a domineering person the same as he treats you?

"?" (Personality)

77. Are people sometimes successful in taking advantage of you?
95. Would you "have it out" with a person who spread untrue rumors about you?
98. Do you prefer a play to a dance?
116. Are you considered to be critical of other people?
120. Do you feel that marriage is essential to your present or future happiness?

These items were typed on belts of adding-machine paper, for attachment to the exposure apparatus which will be described presently. For the first individual experiment three belts were made, containing the same 40 questions, but in different orders. In belt 1, the 20 personality items came first, followed by the 20 achievement items. In belt 4 the order of items in each part was reversed, and the achievement items were placed before the personality items. In belt 5, both parts mixed the types at random, except that wherever the first part of the belt had a personality item, the second part had an achievement item in the corresponding position.

Two additional belts were constructed for use in the second individual experiment. The same items were used again, but in still different arrangements. Belt 2n⁴ began with 10 achievement items. This part was followed by 10 personality items. In the third

part the items were mixed. Part 4 began with 5 personality items and finished with 5 achievement items. Belt 3n used the same general scheme, except that wherever belt 2n had personality items, belt 3n had achievement items in the corresponding positions. Besides alternating the positions of the two types of items in belts 2n and 3n, all items appearing in either of the first two parts of belt 2n were shifted to the last two parts of belt 3n, so that the effects of serial position (expected to be a factor influencing both the psychogalvanic responses and the memory scores) might be minimized as far as possible.

Lists showing the order of the items in each belt are given in the Appendix.

III. Apparatus for the Individual Experiments.

Because of financial limitations, the experimenter was forced to carry on the individual testing without any sort of assistance. It was therefore necessary to create a mechanical setup which would make it possible for one person to secure simultaneous records of reaction time, psychogalvanic deflections, and the subject's answers.

The complete machinery about to be described was used in the first series of individual tests (experiment 1). Certain defects, however, developed during the earlier part of the work, and a portion of the complicated machinery was abandoned in favor of simpler methods of recording, as will be explained later.

The Room. The individual tests were given in an inner room about 10 feet square, with cement floor and plaster walls. The room was not sound-proof, but was sufficiently well isolated from the rest of the building to be free from external noise most of the time. Figure 1 is a floor-diagram of the experimental room, and indicates the positions of the major divisions of the apparatus used in the individual testing.

The central features of the apparatus will now be briefly described.

The Polygraph Recorder. The polygraph recorder, constructed originally for recording reaction times, psychogalvanic responses, and the subject's answers on a single continuous record sheet, was built around a level polished steel writing surface about 6 x 10 inches in size, over which a spring motor driven kymograph on one side of the writing surface pulled a strip of paper 10 inches wide from a supply roll on the other side, at a slow

⁴Other belts corresponding to the missing numbers (2, 3 and 6) were planned or constructed but were not used. A belt numbered 1n was used on 3 subjects between experiments 1 and 2, and will be mentioned briefly later. The original numbers of the belts have been retained in this discussion for sake of convenience.

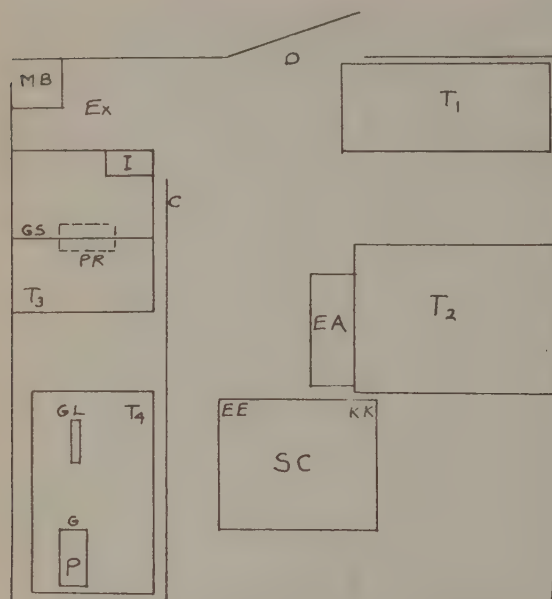


Figure 1. Floor Plan of the Experimental Room.

CODE

- MB--Metronome box, sound-reducing
 D--Door
 T₁--Utility table
 Ex--Experimenter's station during the testing
 I--Instruments for controlling the apparatus
 C--Curtain for concealing the apparatus
 T₃--Double-deck table
 GS--Glass scale, for deflection readings
 PR(dotted)--Polygraph recorder, on lower deck
 T₄--Table for Psychogalvanograph
 P--Wechsler Psychogalvanograph
 G--Galvanometer
 GL--Galvanometer light
 SG--Subject's chair
 EE and KK--Electrodes and answer keys
 EA--Exposure apparatus, lowered into place
 T₂--Table holding exposure apparatus

rate. Four small fountain pens, adjusted to write at touch, were suspended at a natural writing angle with their points resting on the paper in a straight line perpendicular to the direction of motion of the paper. The suspensions were adjustable, so that the alignment of the pens could be tested and re-adjusted for each subject. A small weight was hung from each pen in order to increase the firmness of contact with the paper and thus promote a free flow of ink. Each pen was connected with the moving arm of a small electromagnet, so that when the circuit through the electromagnet was closed the

point of the pen was displaced slightly, thus making a jog in the otherwise straight line traced by the pen on the moving paper. Each electromagnet (except the one connected with the time marker) was also in parallel circuit with a colored bulb which flashed whenever the circuit was closed, thus informing the experimenter that a corresponding record had been made on the moving paper.

A square metal bar, placed above and in a horizontal line directly over the pen points, supported a small metal carriage sliding back and forth along the bar and carrying a pencil-lead holder which, by means of a light spring, kept the lead in contact with the paper.

Two of the pens recorded respectively the "Yes" and "No" answers, and were operated by telegraph keys placed at the end of the right arm of the subject's chair. A third pen was operated by a make-and-break contact on the exposure apparatus, the circuit being closed momentarily each time a new stimulus was presented. The fourth pen was connected in series with a mercury-contact metronome beating at half-second intervals. The metronome was enclosed in a padded box, to deaden the sound. Current for the colored bulbs and for the electromagnets was furnished by an ordinary 6-volt automobile storage battery.

The sliding carriage was connected, by means of a flexible wire cable and pulleys, with a pointer which moved back and forth horizontally on a metal bar placed directly in front of the frosted glass scale which formed a part of the psychogalvanic response recorder, next to be described.

The polygraph recorder, kymograph, paper supply roll and storage battery rested on a firm platform about 3 feet square, four inches above the floor level. This platform was a sort of lower deck for the table (T₃, Figure 1) which was built over and above it to hold the PGR scale and the various control instruments for the electrical circuits. The entire platform was illuminated by means of a 7.5 watt bulb in a socket affixed to the under side of the upper deck of the table, and a convenient aperture cut through the top deck enabled the experimenter to glance down occasionally and thus make sure that the polygraph was functioning smoothly.

Figure 2, diagrams the electrical circuits employed in the operation of the polygraph recorder.

The Psychogalvanic Response Recorder. This portion of the apparatus employed a Wechsler Psychogalvanograph, of the type manufactured by the C. H. Stoelting Co., and certain accessories.

The psychogalvanograph in its original

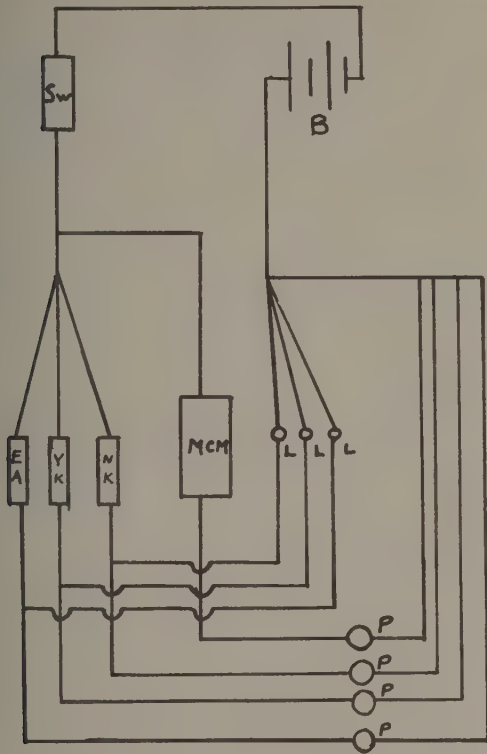


Figure 2. Electrical Circuits for the Polygraph Recorder.

CODE

- Sw---Switch
- B---Storage battery
- EA---Contact on exposure apparatus
- YK and NK---Answer keys
- MCM---Mercury contact metronome
- LLL---Colored lights
- PPPP---Recording pens and electromagnets

form consists of a dark-chamber box about 17 x 5 x 12 inches in size, with a d'Arsonval galvanometer at one end. A special flash-light shaped-device inside the box is so arranged that the light from a single straight-filament bulb is captured by a narrow slit-like opening and is focussed as a vertical hair line on the mirror of the galvanometer. From there it is reflected to the other end of the dark-chamber, where it is caught by a translucent millimeter scale which may be read from the outside, and is also thrown upon a slit behind which a photographic film may be pulled slowly by a small motor attached to the exterior of the box, thus making it possible to photograph the deflections if desired. The top of the box contains a Wheatstone bridge with fixed resistances in two arms, calibrat-

ed rheostats for changing the resistance in the third arm, and connections for placing the subject (via the electrodes) in the fourth arm. A double switch makes it possible to use either of two circuits. The first places the subject in simple series with the galvanometer and source of the exsomatic current (furnished by dry batteries within the box). The second circuit places the subject in the Wheatstone bridge, where his resistance may be balanced and measured. By means of shunts (also located on the box top) the sensitivity of the galvanometer (and therefore the amplitude of the deflections) may be reduced to $\frac{1}{2}$, $\frac{1}{4}$, etc. Connections are also provided for a stimulus key and a timing device, each supposed to make its record by means of a tiny light flashing behind a pin-point aperture near the margin of the moving film. A transformer within the box steps down the ordinary 110-volt lighting current to voltages suitable for the lights mentioned above. The photographic portion of the apparatus was not used at all in this investigation.

The Psychogalvanograph stood entirely apart from the rest of the apparatus, on a separate support (T_4 , Figure 1), which had thick felt squares placed under the legs to help reduce vibrations coming from the building. The parts of the original instrument were rearranged so that the hair line of light reflected from the mirror of the galvanometer was thrown, not upon the translucent scale provided in the instrument itself, but upon a larger frosted glass scale 32 inches long placed on a separate table (T_3 , Figure 1) at a distance of about 63 inches from the galvanometer, thus increasing the amplitude of the deflections about 4 times. The same table also held the Wheatstone bridge top, which had been removed from the Psychogalvanograph, a potentiometer (200,000 ohms), a voltmeter (15 volt Weston, reading in quarter-volts), and three $4\frac{1}{2}$ volt Eveready C batteries which furnished the source of the exsomatic current, besides the colored bulbs and switches for controlling the various circuits. The external battery and potentiometer circuit used to maintain a constant intensity of current in these experiments is shown in Figure 3.

As the beam of light reflected from the galvanometer mirror moved along the frosted glass scale in response to skin resistance changes in the subject, the experimenter followed the motion of the beam by means of the pointer already described, this motion being transferred to the pencil lead, thus tracing on the moving paper a curve substantially equivalent to the photographic record which the machine is supposed to make.

The tables which held the Psychogal-

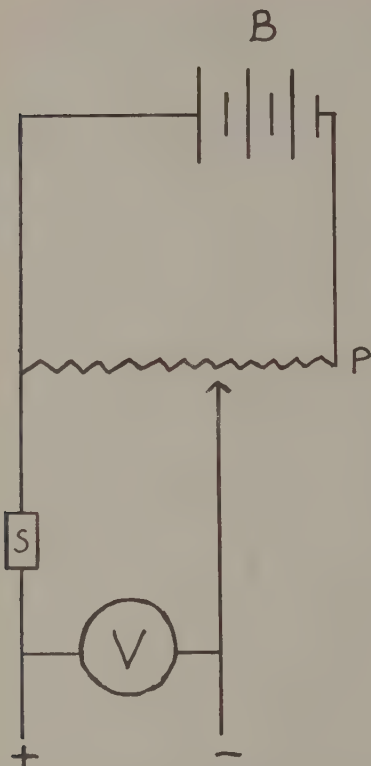


Figure 3. External Battery Circuit for the Psychogalvanograph.

CODE

B--"C" Batteries
 P--Potentiometer
 V--Voltmeter
 S--Switch
 + - --Battery contacts on the Psychogalvanograph

vanograph and the recording instruments were both hidden from the view of the subject during an experiment by a black curtain (C, Figure 1). Before the experiment began, the curtain was lifted while the necessary explanations and instructions were being given to the subject.

The Exposure Apparatus. The subject sat in a comfortable upholstered Morris-chair. The left arm was supported by small pillows, and the palm of the hand rested on a felt covered platform about 4 inches square which could be raised or lowered so as to permit

the fingers to dip into the cups of the electrodes⁵ the proper distance without any muscular strain to hold them there. Within easy reach of the right hand were the answer keys.

Directly in front of the subject was the exposure apparatus, hinged to a table so that it could be lifted up out of the way when not in use (EA, Figure 1). The exposure apparatus consisted essentially of a large drum turning by means of a sort of clock weight connected to a cord wound around a smaller drum attached to the main drum. A catch and release mechanism allowed the drum to be turned one space at a time, each space being exposed as long as might be desired. This mechanism was controlled by means of a wire cord extending over pulleys to a point within reach of the experimenter's left hand.

The drum provided 46 spaces, each $1 \times 3\frac{1}{2}$ inches, appearing successively behind an aperture in a black metal shield which hid the rest of the exposure apparatus from the seated subject. The belts mentioned earlier were made to fit this drum. The make-and-break contact operating one of the recording pens was accomplished by means of small brushes which were struck by projecting brass bolt heads each time the drum moved. A small lamp was placed so as to illuminate the aperture clearly, but without unpleasant glare for the subject.

The experimenter also had a small shaded lamp illuminating the Wheatstone bridge and other instruments. His right hand traced the beam of light with the pointer (in the first experiment), the left hand being free to manipulate the bridge rhostats, potentiometer, switches, exposure apparatus cord, etc., as desired.

Remarks. The exposure apparatus, like the polygraph recorder, was specially designed by the writer for use in this experiment, after finding ready-made devices either unsatisfactory or too expensive for the straightened budget by which the study had to be financed.

The Psychogalvanograph could not be used in its original form, since it was necessary to take a larger number of records than the regular mechanism allows. The method used for the recording of deflections requires a semi-dark room. The semi-darkness was not believed to be a disturbing factor, since the subject's lamp illuminated his side of the curtain enough to dispel the gloom

⁵Liquid electrodes, also made by the C. H. Stoelting Co., were used. These electrodes will be described more fully on a later page in connection with the discussion of the technique employed for testing the subjects during the first individual experiment.

considerably, and also, the reason for extinguishing the overhead light was carefully explained to each subject. The same precaution was taken with regard to the slight noise incidental to the operation of the recording instruments.

The flexible wire cable used with the pointer was made of material taken from fish-leaders. This cable is small and very flexible and does not stretch like cord. Since the motion of the pointer was cut down to about one sixth by means of a reducing wheel before being transmitted to the sliding carriage, it was desirable to eliminate elasticity as much as possible. However, in spite of the precautions taken to avoid it, there still remained enough inertia in the machine to interfere seriously with the recording of the smaller deflections, and this portion of the apparatus was abandoned at the conclusion of the first experiment.

Measurement of the Psychogalvanic Response. Wechsler's technique for measuring the psychogalvanic response, which was used in this investigation, attempts to compensate for differences and variations in bodily resistance by keeping the intensity of the current (that is, the ratio between the applied e.m.f. and the subject's apparent resistance) approximately constant during the course of an experiment, and also constant for different subjects in the same experiment, thus making it possible to compare their deflections. The subject is first placed in simple series with the galvanometer and the battery, and the potentiometer is manipulated until the desired intensity is reached. This state of affairs is best determined by means of a sensitive milliammeter in the circuit. Such an instrument was not available for use in the present experiment, and an approximate method described by Sears⁶ was used instead. A fixed resistance of 10,000 ohms is placed in the circuit instead of the subject, against an e.m.f. of 1 volt, and notice is taken of the point on the scale to which the beam of light is deflected under these conditions. Then when the intensity is to be adjusted for any subject, the potentiometer is manipulated until the beam of light comes to this same point on the scale, regardless of the magnitude of the initial deflection given by the subject when the circuit is first closed.

While the initial intensity is being adjusted, the galvanometer is shunted to its

lowest sensitivity. After the adjustment has been satisfactorily made, the double switch is turned so as to throw the subject into the Wheatstone bridge, and the galvanometer is shunted to a higher sensitivity. The rheostats are then adjusted until the beam of light returns to whatever point on the scale was originally chosen as a zero point. The resistance of the subject has now been balanced, and may be read directly from the dials of the rheostats. Also, all is in readiness to begin the presentation of the stimuli.

If, as a result of an extra large deflection, the steady "drift" frequently encountered, or some other cause, the resistance level of the subject changes markedly during the course of the experiment, the beam of light will fail to return to the zero point. This indicates that the intensity of the current flowing through the circuit has changed enough to make an adjustment necessary. The adjustment is simply a repetition of the initial procedure; that is, the subject is switched back momentarily into the simple series circuit with the galvanometer shunted to its lowest sensitivity, the intensity is adjusted by means of the potentiometer as at the beginning of the experiment, and the bridge is rebalanced before continuing with the stimuli.

The number of times that it is necessary to make this sort of adjustment varies from subject to subject, and also, obviously, with the narrowness of the limits within which it is desired to keep the intensity constant. Seldom does the beam return exactly to the zero point after a deflection. The most common phenomenon is a steady drift of the spot at which the beam stops, in the direction of decreased resistance. Sometimes, on the contrary, the beam will show a similar steady drift in the direction of higher resistance. A few subjects show both tendencies alternately during the course of an experiment.

In the present experiment, the scale was calibrated in units from +160 to -30. The zero point was taken as 100 on this scale, and before each test the beam of light was adjusted to fall at this point. (A thumbscrew device is provided on the top of the galvanometer for this purpose). As a general rule, readjustments were made during the course of a test whenever the beam failed to return after a deflection to a point between 70 and 130 on the scale. Any attempt to control the

⁶Sears (19) used the same Psychogalvanograph employed in this experiment. The present writer is indebted to Dr. Sears' dissertation, or to his personal advice, for a number of helpful suggestions regarding the technique of pgr measurements and the use of the Psychogalvanograph.

intensity within much narrower limits would have necessitated so many readjustments that the sittings would have been nearly doubled in length.

The theory underlying the Wheatstone bridge circuit and the foregoing method of compensating for changes in resistance, the evidence favoring the use of liquid electrodes, and an account of the construction of the original Psychogalvanograph are contained in Wechsler's monograph (24).

IV. Technique of the First Individual Experiment.

Twenty subjects were tested in the first experiment. An appointed time was set when each subject reserved a two-hour period for the test. This entire length of time was not needed for the test itself, which as a rule required only about an hour including the time necessary to prepare and instruct the subject. The extra hour allowed a generous margin of safety, especially in the cases of those subjects who were late in coming, and also provided the time needed for readjustments of the apparatus before the next subject arrived. Most of the appointments were made for the hours of 8 to 10 or 10 to 12 a.m., or for 1 to 3 or 3 to 5 p.m. No evening tests were given. The maximum number of subjects tested in any single day was four.

Upon arriving for the test, the subject was immediately seated in the chair, and the left hand was prepared for the electrodes. These were of the type designed by Wechsler and claimed by him to be relatively non-polarizable. Each electrode consisted of a porous clay cup of normal NaCl solution resting in a slightly larger glass cup of saturated $ZnSO_4$ solution. Between the two cups was a strip of amalgamated zinc, to which was attached the wire leading to the galvanometer and bridge circuits. The first and third fingers of the left hand were immersed in the cups of salt solution as far as the first joints. Following the method of finger preparation described by Sears,⁷ the ends of the fingers were first carefully washed with alcohol. A coating of vaseline was then placed around the finger above the joint, and over the vaseline, to prevent it from being rubbed off, a strip of $\frac{1}{2}$ inch adhesive tape was wound firmly but not tightly. The platform on which the palm of the hand rested could be raised or lowered by means of supporting bolts, so that the level of the salt solution would come about midway on the tape. Rubber cots were placed on the other fingers, which were allowed to drop in any convenient position

comfortable to the subject. The forearm was supported by pillows suitably placed.

While the fingers were being prepared for placement in the electrodes, the experimenter carefully explained the general nature of the experiment, in an attempt to set the subject at ease as much as possible. Such knowledge as was possessed by most of the subjects connected the psychogalvanic response rather vaguely with collapsing chairs and missing top steps, and so the explanations and assurances that nothing of a startling nature would occur were necessary in order to induce a fairly natural frame of mind.

After the fingers had been placed in the electrodes, the current was turned on and allowed to flow for several minutes before balancing the bridge, in order that the initial changes due to polarization and other factors might become completed. During this time the potentiometer would usually have to be readjusted several times to compensate for the rapid decrease in resistance that ordinarily occurred during the first few minutes that the subject was in the circuit. Meanwhile, the subject was given the necessary instructions. They were printed on a sheet of paper affixed to the front shield of the exposure apparatus, which was lowered into place in front of the subject at this point. The subject's lamp was also switched on. The instructions are given below. In addition to those printed, it was also suggested to the subject that glancing down at the keys (each of which was labelled with its corresponding answer) when pressing them would help greatly to avoid mistakes.

Instructions

Press the RIGHT-HAND key for "YES".

Press the LEFT-HAND key for "NO".

Press BOTH keys at once for "?".

Press ONCE if you find the question relatively EASY to answer.

Press TWICE if you find the question relatively DIFFICULT to answer.

PRESS FIRMLY, holding the key down for about one second or more.

PLEASE TRY NOT TO MOVE, COUGH, OR BREATHE DEEPLY.

DO NOT REMOVE YOUR FINGERS FROM THE CUPS UNTIL TOLD TO DO SO.

TELL THE EXPERIMENTER AT ONCE WHEN THE BLANK SPACE APPEARS.

⁷See footnote 6.

You may change your answer, if you wish to do so, provided you do it before the next question appears. In such cases, the second answer (or the last) will be the one counted.

The directions were read aloud and explained, including practice with the keys and with the changing of spaces in the exposure apparatus, until the subject appeared to understand thoroughly. The subject was warned that the questions would not come at regular intervals, but that he should answer each one as soon as possible and then wait quietly for the next.

Meanwhile, the experimenter had set the recording apparatus in motion, in order to accustom the subject to the slight noise, and had extinguished the overhead light, after explaining that the semi-darkness was necessary for seeing the beam of light clearly.

When all else was ready, the subject was switched into the Wheatstone bridge circuit, the initial resistance was measured and recorded, and the series of questions followed. After the answer had been given for an item, sufficient time was allowed for the beam of light to come to rest if there had been a deflection. If no deflection had occurred at the end of 15 or 20 seconds, the next question was presented without further waiting.

When the blank space appeared, at the end of the first series of 20 questions, the current was turned off, the recording instruments stopped, the upper light switched on, and the subject was given pencil and paper with instructions to write down as many of the questions as possible without puzzling over them a long time. The idea was to get a record of simply those questions which had made enough impression to be immediately "on tap", as it were. As soon as the subject became obviously "stuck", the paper was taken away. (While the subject was writing the experimenter had inspected the pens, wound the metronome, made necessary notations on the record sheet, etc.) The current was then turned on again, and the whole procedure, including a second recall period, was repeated with the second list of 20 items.

Before the subject left the room, he was requested not to tell any other members of the class about the questions that were being used or the fact that recall was demanded after each set.

V. Technique of the Second Individual Experiment.

Before many subjects had taken the individual test as just described, certain

weaknesses in the method became apparent. (The trouble with the pointer has been mentioned. Other difficulties will be discussed in connection with the treatment of the results.)

During the first experiment, belts 1, 4 and 5 had been used in rotation with successive subjects. Also, alternate subjects with each belt were requested to speak the answer aloud at the instant of pressing the key. This was in the interests of a little side-investigation concerning the effects of vocalization on the magnitude of the deflections. Thus half of the records represented "vocal" answers, and half represented "non-vocal" answers.

After 20 subjects had been tested, certain changes were made in the technique of the experiment. These changes constitute the points differentiating experiment 2 from experiment 1.

First, after a short trial period with another order of presentation (belt 1n) of the questions, belts 2n and 3n were constructed and used alternately with the remaining subjects.

Second, there were no long breaks in the series of questions comparable with the interval between the two parts of experiment 1. The only breaks in the continuous presentation of the questions were blank spaces preceding each of the four parts, and these were used simply for the purpose of informing the subject of the nature of the items in the next part. The warnings for the first and second parts were, "The first (or next) part will consist of personality (or information) test questions." The warning for the third part informed the subject that the questions would be "mixed, partly from one test and partly from the other." The warning for the fourth part named the kind of items which came first, but said nothing about the change to the other type at the sixth item.

Third, the elimination of the long break made it necessary to substitute one single recall period, after all of the questions had been answered, for the double recall period used in experiment 1.

Fourth, instead of tracing the moving beam of light with the pointer, the experimenter simply observed the extent of the excursions and recorded the pertinent data on specially prepared record sheets. This method was found to be entirely feasible, since the beam of light normally does not move very rapidly, and the scale numbers were large enough to be read easily and accurately. However, this method had the disadvantage of precluding the possibility of measuring the latent periods of the deflections, measurements

which were originally intended to be included in the data.

In connection with this matter of reading the magnitudes of the deflections, a word is in order concerning unwanted deflections due to movement or some other chance extraneous cause. As various writers on the subject have observed, one does not have to work with the psychogalvanic response very long before he becomes able to detect these extraneous deflections with considerable assurance. They have no latent periods, they occur mostly during the time when the beam of light would ordinarily be at rest while the subject is awaiting the next stimulus, they are more rapid than the regular deflections (especially on the return journey), and they possess other distinctive characteristics which are not easily described but are fairly easily detected. In the present investigation, all such suspicious deflections were simply ignored and not recorded at all.

With the exception of the four alterations noted above, the main outlines of the experimental technique, including the mechanical details, remained unchanged for experiment 2. The practice of alternating vocal and non-vocal answers was also retained. Most of the important results presented later are based on the second experiment.

VI. Interpretation of the Records.

Figure 4 reproduces a sample portion of one of the record sheets from experiment 1. The wavy line at the top is the pgr record, the falls in the curve representing deflections away from the zero point, which was at 100 on the glass scale. The second line is the time line, each jog representing one half second. The arrow on the time line indicates the direction of motion of the paper. The record is therefore read from right to left. The third line is the stimulus line, a jog being made therein whenever a new question was presented. The fourth and last lines are respectively indicative of "Yes" and "No" as answered by the subject's pressure on the keys. A judgment of "difficult" would be indicated by two jogs in immediate succession on the line or lines corresponding to the answer.

The four recording pens and the pgr pencil point were kept in a straight line perpendicular to the direction of motion of the paper. When transcribing the data, perpendicular lines were drawn with a square so as to cut all five lines on the record sheet. The time intervals could then be read by counting the jogs on the time line between the perpendiculars corresponding to stimulus and answer. The pgr deflections were also read by measuring the distance along perpendiculars between the highest and lowest points on the curve following each stimulus.

In the second experiment, obviously, this same record sheet held only the four lower lines, the actual magnitudes of the deflections having been recorded elsewhere.

In both experiments, all data for each subject were placed on two graph sheets. Figure 5 reproduces one such sheet from experiment 2. The deflection curve has been diagrammed at the right, the numbers at the top corresponding to the scale divisions. The numbers at the extreme right mean: subject number 32, woman, belt 3n, vocal answers.

The letter R adjacent to a number on the pgr line indicates that the corresponding item was recalled in the memory test following the answering of the questions.

The first column of figures toward the left margin records the magnitudes of the deflections, then come the reaction times in the next column, and finally the answers. Time intervals were measured to the nearest half second. A second reaction time and answer in parentheses indicate that the answer was changed that many seconds after the original answer. In such cases the latter answer was the one counted, and the total time elapsing before that answer was given was reckoned as the reaction time. "D" at the left of an answer indicates that the subject rated that question as relatively difficult. The dotted line indicates a readjustment of the intensity of the current to compensate for change of resistance level in the subject.

Reference to the photographs on the following pages may help to clear up some of the apparatus descriptions which have been given.

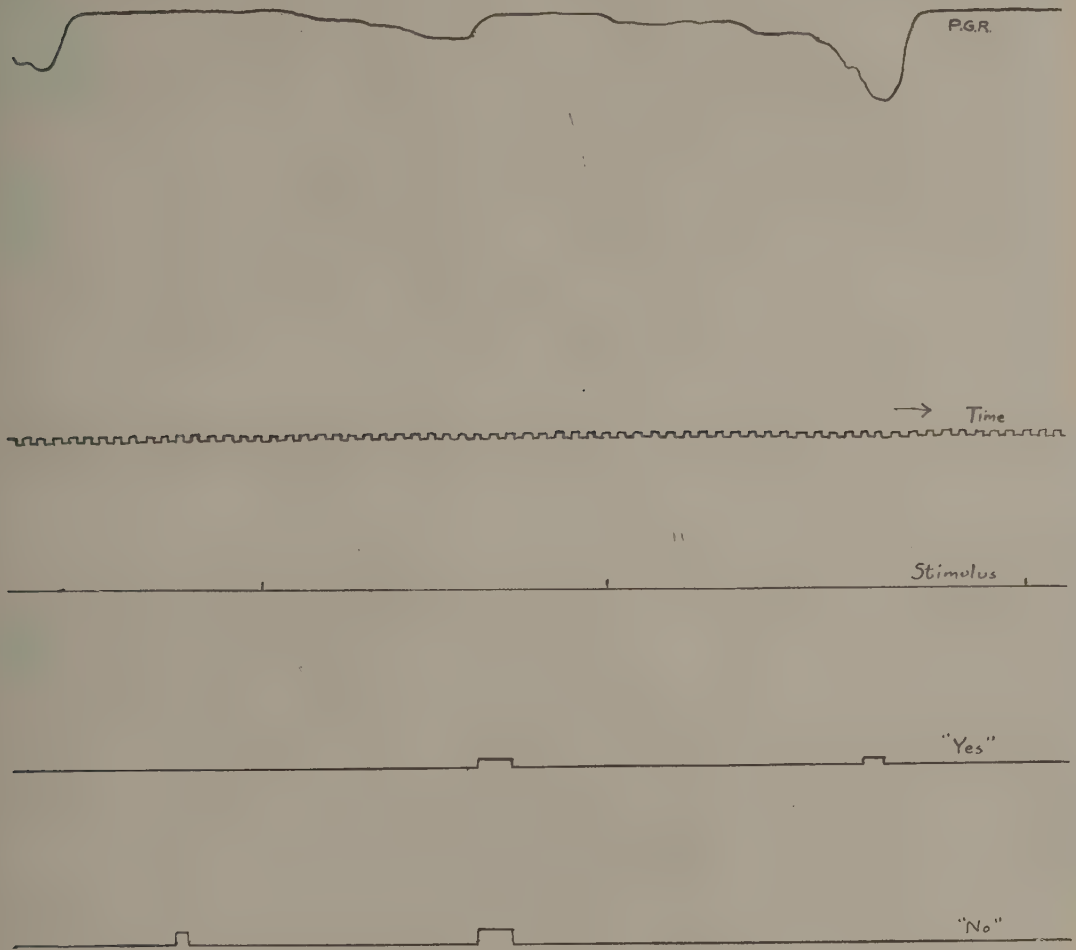


Figure 4. Sample Portion of a Record Sheet
from the First Experiment.

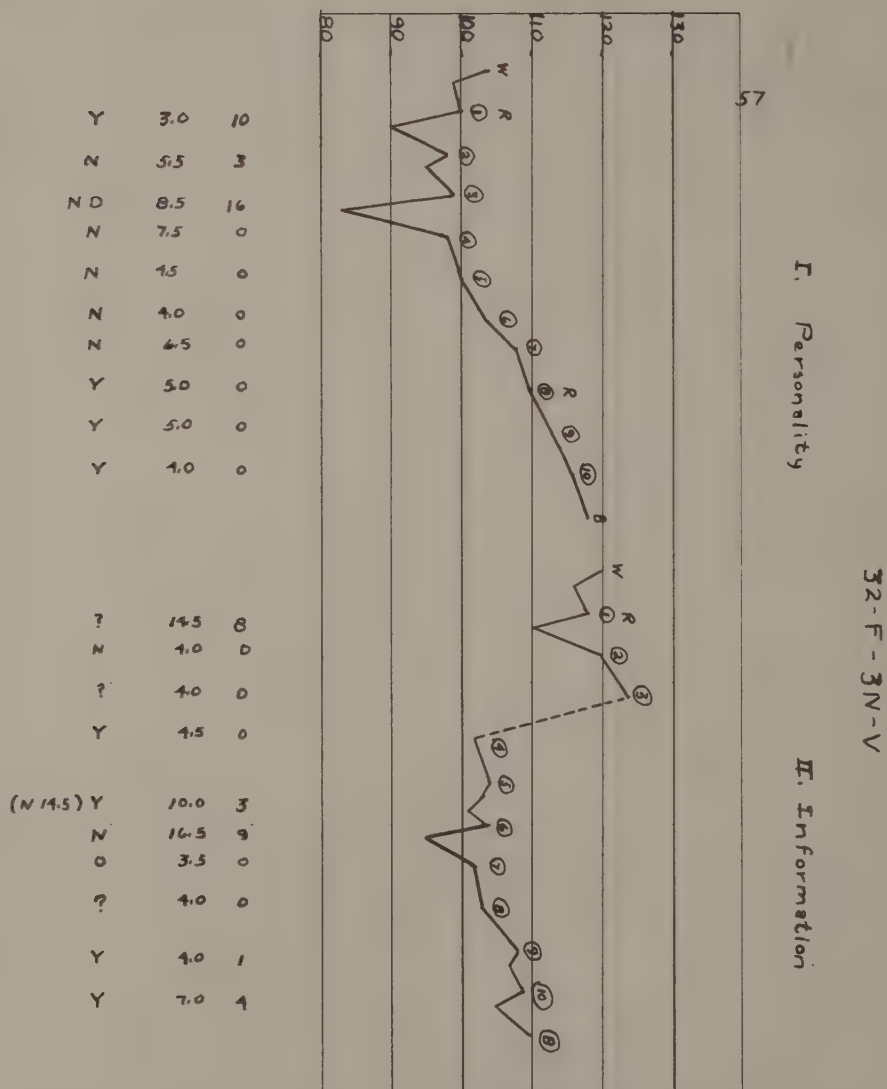
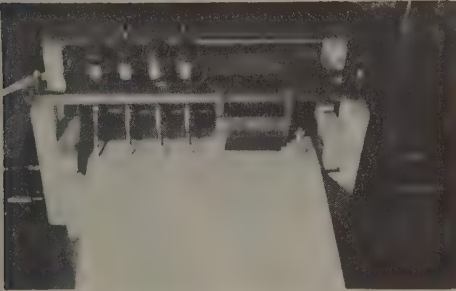
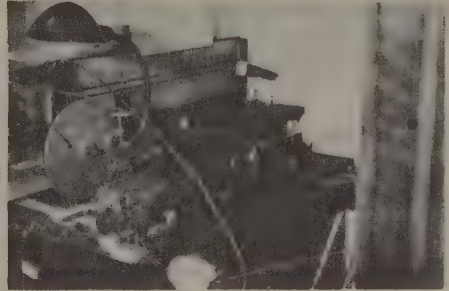


Figure 5. Sample Record Graph from the Second Experiment.



Close-up view of the polygraph recorder, showing the pens and the sliding pencil-point carriage

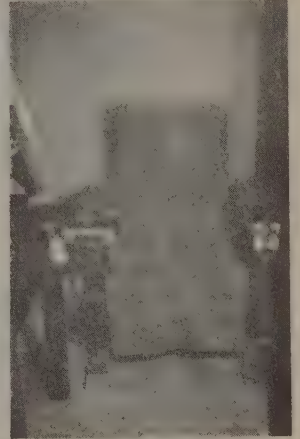


Upper deck of Table T₃, showing the glass scale, sliding pointer, colored bulbs, and control instruments for operating the Psychogalvanograph

The wire draped behind the lamp does not belong there



Lower deck of Table T₃, showing the polygraph recorder, kymograph, paper supply roll, battery and metronome box

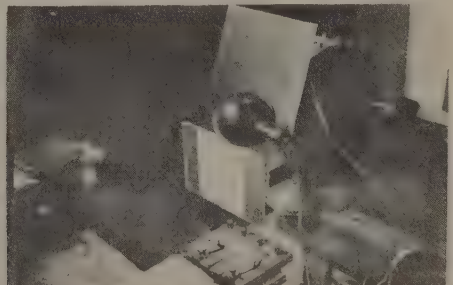


(a) Subject's Chair (b)

With exposure apparatus in place, and (b) with exposure apparatus lifted out of the way



The Wechsler Psychogalvanograph as used in this experiment



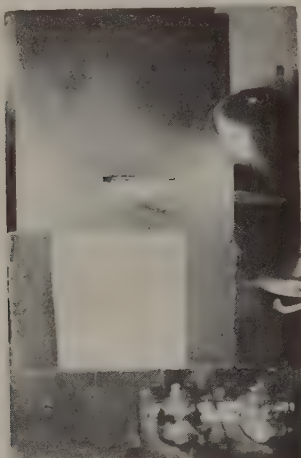
Exposure apparatus, subject's lamp, answers keys, and electrodes: ensemble view



The electrodes and platform for supporting the left hand, with hand in place



The experimenter in working position for Experiment 2



Subject's view of the exposure apparatus and instruction sheet

The printing on the instruction sheet did not photograph

The hook near the fingertips of the left hand controlled the exposure apparatus. The curtain is lifted above its regular position, in order to show the instruments clearly

Chapter III

RESULTS OF THE GROUP TESTING AND OF THE FIRST INDIVIDUAL EXPERIMENT

I. Results of the Group Testing.

It has already been indicated that the primary purpose of the preliminary group testing was to secure data on which to base the selection of items to be used in the individual testing, and the selected items were listed in Chapter II. A detailed analysis of the scores on the Bernreuter Personality Inventory and the General Information Test would be unprofitable, since the results would have no important bearing on the specific problem of this investigation. However, the data also afford the opportunity to make certain direct comparisons which are pertinent to the problem, and it is with these aspects of the group testing results that the present section will deal.

Scores on the General Information Test. Table I shows the distribution of the scores made on the General Information Test by the 86 subjects whose papers were scored. The score was taken as the number right. The total distribution has a distinct skewness, indicating an excess of smaller scores, but this skewness is not large enough to disturb the balance seriously. The difference between the mean scores for the men and women is only about half of its S.D., hence relatively insignificant.

Table I
DISTRIBUTION OF SCORES OBTAINED ON THE
GENERAL INFORMATION TEST

Interval	Men	Women	Total
100-109	1		1
90- 99	3		6
80- 89	6	11	17
70- 79	5	20	25
60- 69	7	20	27
50- 59	2	6	8
40- 49	2		2
N	26	60	86
Mean	73.42	72.08	72.48
S.D.	15.44	10.02	12.14
V	21.03	13.9	16.75

V, in Table I and in following Tables, is the Pearson Coefficient of Variation (100 S.D./Mean). The men are about 1.51 times as variable as the women in Table I.

Scores on the Bernreuter Personality Inventory. Table II shows the distribution of the percentile scores for scale Bl-N (neurotic tendency). There are several contrasts between this distribution and the distribution of General Information Scores. In the first place, the larger S.D.'s and V's of the Bernreuter distribution indicate that the group is less homogeneous with respect to the Bl-N measures than it is with respect to the General Information measures.

Table II
DISTRIBUTION OF SCORES OBTAINED ON THE
BERNREUTER PERSONALITY INVENTORY

Interval	Men	Women	Total
91-100	3	3	6
81- 90	3	6	9
71- 80	2	7	9
61- 70	4	3	7
51- 60	3	1	4
41- 50	4	5	9
31- 40		10	10
21- 30	1	9	10
11- 20	2	13	15
0- 10	4	3	7
N	26	60	86
Mean	53.73	43.35	46.48
S.D.	29.62	27.32	28.57
V	55.11	63.02	61.46

In the second place, the Bernreuter distributions for both men and women show a trend toward bimodality, especially the distribution for the women, which is absent from the General Information distributions. Third, in the Bernreuter distribution, the women are 1.14 times as variable as the men, the reverse of the corresponding tendency in the General Information distribution. Lastly, there is a much more significant difference between the Bernreuter means for the two sexes than between

the General Information means. (The corresponding Diff./S.D. diff ratios¹ are 1.53 and .40.)

These differences between the Bernreuter scores and the General Information scores for the same subjects might be interpreted from either of two general viewpoints: either (1) the Bernreuter scores are measuring something which actually differs in the indicated ways from that which is measured by the General Information scores, or (2) the different subjective conditions supposedly involved in the Bernreuter testing so affect the responses of the subjects that the resulting scores are distributed differently from the General Information scores regardless of whether the Bernreuter test is truly measuring anything definite. The second assumption would be obviously in harmony with the criticisms cited in Chapter I; for if the score on such a test as the Bernreuter actually is determined more by the whim of the subject than by the real facts regarding the trait which is supposedly being measured, then it would be reasonable to expect all sorts of vagaries and distortions in a distribution of the scores for any considerable number of subjects.

An example will make the above point somewhat clearer. In the Bernreuter distribution, 68% of the women score lower than the median (the 50th percentile) for the population (college women) of which they are supposed to be a part, but only 42% of the men have correspondingly low scores. In harmony with the first assumption, this fact might be interpreted to mean that the men in this group actually are more neurotic on the average than the general run of college men, and that the women actually are less neurotic on the average than the general run of college women. In harmony with the second assumption, the same fact might be interpreted to mean that the women in this group of subjects were more prone to give the answers which would enable them to appear in the most favorable light, while the men were more honest and therefore more truly revealed the actual extent of the neuroticism existing among them. The further fact that the scores for the men, although slightly bimodal, are much more evenly distributed through the various decile intervals than are the scores for the women, would seem to lend support to this second interpretation, since the distributions of per-

centile scores for groups which are typical of the population from which the norms are derived should be approximately rectangular anyway.

Comparison of the Times Required for Completion of the Two Tests. A second phase of the group testing which makes possible a direct comparison of the two testing situations is found in the records of the number of minutes utilized by each subject in answering each of the tests. It was partly because of the desire to make such a comparison that the General Information Test was caused to include the same number of items as the Bernreuter test.

Figure 6 and its accompanying tabular insert² show a distinct separation of the curves representing the distributions of times for the two tests. (All 88 subjects are included here.) The shorter times predominate in the Bernreuter distribution up to the 12-minute point, while the longer times clearly predominate likewise in the General Information distribution beginning with the 15-minute point. That is, the shorter times are found mostly in the Bernreuter distribution, and the longer times in the General Information distribution, indicating a longer average time for the General Information Test. Table III gives the data concerning the magnitude and reliability of the difference. Table III is typical of the many similar tables which will be presented later. Each Table will state (1) the average measures (means) which are being compared, (2) the S.D.'s of the distributions, (3) the Coefficients of Variation, (4) the S.D.'s of the averages, (5) the difference between the averages, (6) the S.D. of the difference, (7) the correlation, if any, and (8) the Critical Ratio, together with its interpretation in terms of the number of chances in 100 that the difference is a real difference greater than zero, the latter taken from Garrett's table (8, page 134).

Table III
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN (1) THE AVERAGE NUMBER OF MINUTES USED IN ANSWERING THE ITEMS OF THE BERNREUTER PERSONALITY INVENTORY, AND (2) THE AVERAGE NUMBER OF MINUTES USED IN ANSWERING THE ITEMS OF THE GENERAL INFORMATION TEST. (88 Cases)

	S.D.	S.D.	S.D.			
	Average (dis)	V	(aver)	Diff. (diff)	r	C.R.
(1)	12.97	3.16	24.36	.33	2.28	.31 .50 7.35
(2)	15.25	2.81	18.46	.30		(100)

¹Hereafter the abbreviation C.R. (critical ratio) will be used to indicate the Diff./S.D.diff. ratio used in estimating the significance of an obtained difference between two averages.

²Each Table presented in this study will be accompanied by an untitled tabular insert of this sort giving the numerical data used in the construction of the Figure.

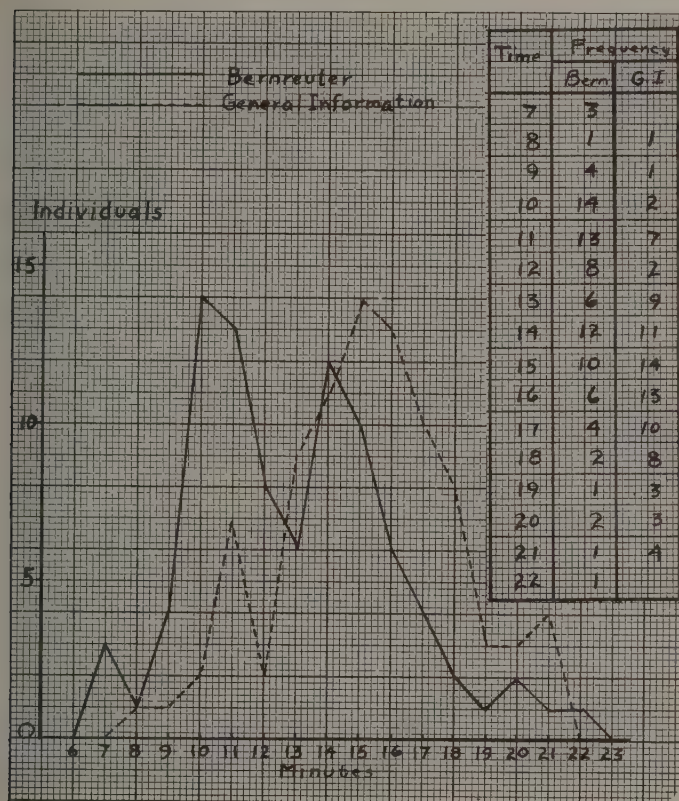


Figure 6. Times Required for Completion of the Bernreuter Personality Inventory and the General Information Test.

The size of the C.R. indicates complete reliability for this difference. The individual records show that 58 subjects agree with the group tendency, 18 used the same number of minutes for each test, and only 12 required a longer time period for the Bernreuter test. It therefore seems reasonable to conclude that, under standard group-testing conditions, the personality items are answered more rapidly on the average than the achievement items. The Bernreuter times also are more variable than the General Information times, the ratio being 1.32.

The high reliability of the difference in Table III is of interest in con-

nection with the reaction time findings of the individual experiments, wherein a similar but much less reliable difference is found between the reaction times for the selected personality and achievement items.

Comparison of the Number of "?"

Answers to the Items of the Two Tests. As has already been stated, a "?" alternative, not commonly found in connection with the true-false item, was added in this case in order that the external forms of the two types of questions might be as much alike as possible. In view of this fact, it was thought that it might be of some value to make a comparison of the relative frequencies with which the "?" answer was used on the two tests. There is little reason to doubt that the "?" answers given to the achievement items represent, at least in a very large majority of the cases, either definite ignorance or honest doubt regarding the facts touched upon by the test items.³ In the case of the personality items, however, the interpretation to be placed on the "?" answer is by no means as clear as in the case of the achievement items. A "?" answer to a personality item may represent actual ignorance or honest doubt, but it may also represent an attempted evasion of the issue, a subterfuge adopted when the subject does not wish to commit himself either way. It would therefore be quite in harmony with the criticisms of the personality test to expect a rather frequent usage of the "?" answer in such tests as the Bernreuter, that being perhaps one of the easiest ways to resolve the conflict of subjective tendencies, granting for the sake of argument that such conflict does occur frequently.⁴

The actual facts as revealed by a comparison of the "?" answers given by the 88 subjects in the two tests are at complete variance with the arguments just advanced. Figure 7, on the following page, shows that the distribution of "?" answers for the Bernreuter test was not at all similar to the

³See page 54 for another possibility.

⁴The comments of the subjects frequently lend support to this point of view. Typical comments are: "You can't answer just yes or no"; "It all depends on the circumstances"; "Sometimes I would do one and sometimes the other" or "I was never in that situation"; etc. It would not seem illogical to expect that the "?" answer would appear often under such circumstances as these. Indeed, the provision for only a flat Yes or No answer was considered one of the weaknesses of the original Psychoneurotic Inventory, and later inventories have sometimes provided for graded answers such as "Always", "Often", "Sometimes", "Seldom", "Never", etc.

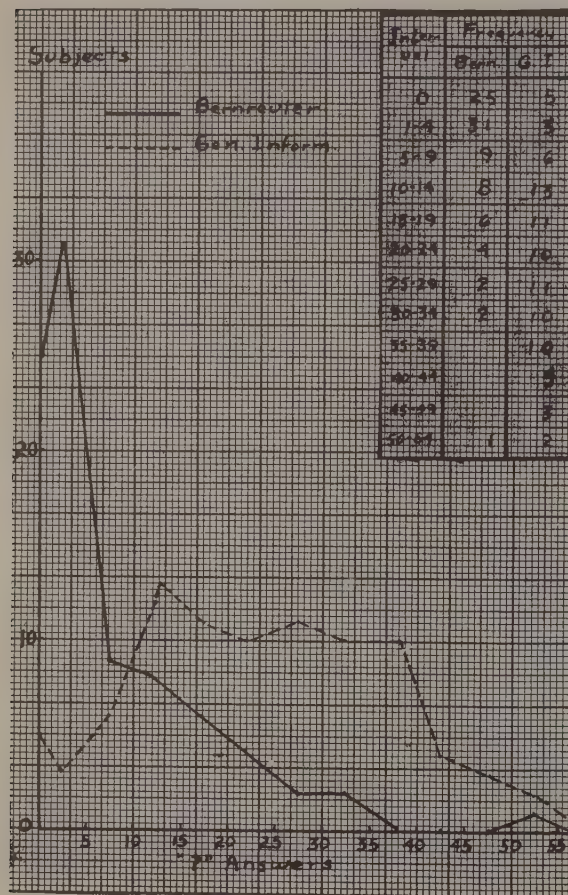


Figure 7. Frequency of "?" Answers Given by 88 Subjects to the Items of the Bernreuter Personality Inventory and the General Information Test.

corresponding distribution for the General Information Test. The ranges are about the same, but there is no other common characteristic of the two distributions. The distribution for the General Information Test approximates a rough symmetry, but the Bernreuter distribution is skewed beyond any pretense of symmetry. Nearly 30% of the subjects did not use the "?" answer at all in the Bernreuter test, while only one-fifth as many fall into the same category in the General Information Test. At the other end of the scale 40 subjects employ the "?" answer 25 or more times in the General Information Test, while only one-eighth as many do likewise in the Bernreuter test.

Table IV presents the reliability data concerning the difference under discussion.

Table IV
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN (1) THE AVERAGE NUMBER OF TIMES THE "?" ANSWER WAS USED IN THE BERNREUTER PERSONALITY INVENTORY, AND (2) THE NUMBER OF TIMES THE "?" ANSWER WAS USED IN THE GENERAL INFORMATION TEST. (88 Subjects)

	S.D.	S.D.	S.D.				
Average	(dis)	V	(aver)	Diff.	(diff)	r	C.R.
(1)	6.96	7.98	114.65	.84	16.36	1.60	.08 10.22
(2)	23.32	13.50	57.89	1.43			(100)

In spite of the great difference in the means, the relative variability of the number of "?" answers given to the Bernreuter test is nearly twice as great as for the General Information Test.

Too much significance must not be attached to the high reliability of the difference in Table IV, since the General Information Test admittedly contained a number of items touching matters entirely outside of the possible knowledge of some of the subjects. The point to be emphasized is not the large number of "?" answers given to the General Information items, but rather the small number given to the Bernreuter items. It is certainly fair to conclude that the "?" answer is not used as a loop-hole to any considerable extent in the answering of the Bernreuter items. The subjects may feel in doubt when answering these items, but they tend to take a definite "Yes" or "No" stand despite the doubt.

Comparison of the Number of Changed Answers on the Two Tests. Another possible comparison between the group-test performances, which certainly has some bearing on the subjective conditions surrounding the testing, deals with the number of changed answers appearing on the blanks. The blanks were inspected for the obvious instances in which one answer had been erased or crossed out and another substituted for it. It was found that the average number of changes was slightly larger for the General Information Test. Table V gives the reliability data.

Table V
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN (1) THE AVERAGE NUMBER OF CHANGED ANSWERS ON THE BERNREUTER PERSONALITY INVENTORY, AND (2) THE AVERAGE NUMBER OF CHANGED ANSWERS ON THE GENERAL INFORMATION TEST. (86 Subjects)

	S.D.	S.D.	S.D.				
Average	(dis)	V	(aver)	Diff.	(diff)	r	C.R.
(1)	2.10	1.86	88.6	.20	.37	.19	.52 1.94
(2)	2.47	1.89	76.51	.20			(97)

The number of changed Information answers was larger for 41 subjects, the same for 21, and smaller for 24, in comparison with the number of changed personality answers. The difference between the means is not large enough to be completely reliable, nor is it so small as to be entirely without significance. The respective distributions are shown in Figure 8. The smaller number of changes are somewhat more numerous in the Bernreuter distribution, and vice versa. The relative variability is larger for the Bernreuter changes (ratio, 1.15). The conclusion that the changes tend to be slightly more numerous in the Information test is probably justified, but the difference, if real, is small, and is not consistent with all subjects.

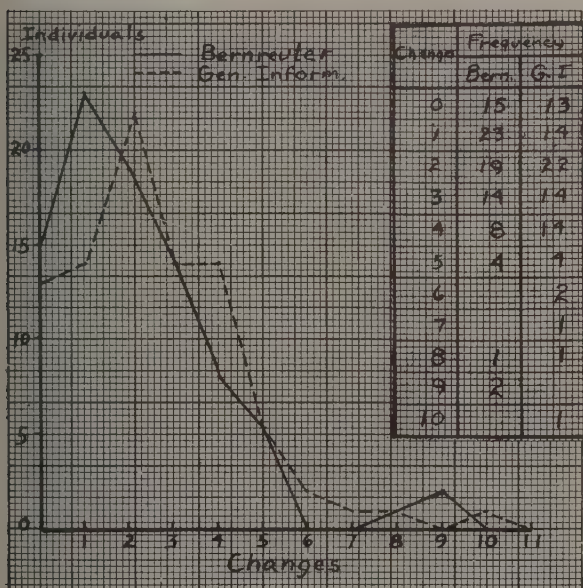


Figure 8. Distribution of Changed Answers in the Bernreuter Personality Inventory and the General Information Test.

Significance of the Group-Testing Results. It has been demonstrated that the Bernreuter scores tend to be distributed bimodally in contrast with the relative symmetry of the General Information score distributions, and that the sex difference is more reliable for the Bernreuter scores. It has been further demonstrated that the average subject uses less time for taking the Bernreuter test than for taking the General Information Test, that the "?" answer is used but very little in answering the Bernreuter items (both absolutely and in contrast with

the other test), and that (probably) there is a tendency for answers once recorded on the Bernreuter blanks to be slightly less subject to changes than the answers once recorded on the General Information blanks. The various Bernreuter measures, in general, are more variable than the corresponding General Information measures.

As was stated in the Introduction, it is not possible to generalize very far on the basis of such results. It must be remarked, however, that the evidence presented thus far hardly pictures the personality testing situation as one of subjective conflict and turmoil. In comparison with the achievement answers, the personality answers are given more rapidly, are more definite, and are more stable once they have been made. The only contrary evidence is the generally greater variability of the personality responses. On the whole, the subjects appear to be working at a higher efficiency level when answering the personality items as compared with the corresponding level for the achievement answers. This fact does not deny the possibility that the apparently greater efficiency may be purchased at the price of greater energy expenditure, a possibility on which the studies of the psychogalvanic responses in a later chapter will throw further light.

II. Results of the First Individual Experiment.

Although it was not so intended in the original plans, that portion of the individual testing which has been designated as the first individual experiment (or experiment 1) was actually relegated to the role of a preliminary try-out investigation. The results therefore will be treated only very briefly, and mostly for the purpose of further clarifying the reasons for the changes in procedure (especially in connection with the psychogalvanic responses) which collectively constituted the difference between experiment 1 and experiment 2. The major conclusions to be drawn from this investigation will be based on the results of the second individual experiment and the group-testing.

Psychogalvanic Responses in Experiment 1. A total of 20 subjects participated in experiment 1, including 3 men and 17 women. Belts 1 and 4 were used by 7 subjects each, and belt 5 by 6 subjects. Vocal answers were given by 9 of the subjects, non-vocal answers by 11. The 20 subjects made a total of 800 responses to the 40 items. Recording defects made necessary the rejection of 24 of the responses. Of the remaining 776 responses, 393 (50.6%) were accompanied by deflections

sufficiently large to be measured on the record sheets. The personality item responses were accompanied by measurable deflections in 53.4% of the cases, the achievement item responses in 47.7% of the cases. The grand average of all deflections was 5.41 scale divisions (approximately equivalent to millimeters on the original scale of the Psychogalvanograph). The mean personality item deflection was 5.89, and the mean achievement item deflection was 4.96. The S.D.'s were not computed, but the evidence from experiment 2 indicates that they would be about equal, numerically, to the averages.

Taken at face value, these data would point to a slight tendency for the personality item responses to be accompanied by measurable deflections more frequently than the achievement item responses, and also to a tendency for the personality deflections⁵ to be larger than the achievement deflections. Certain recording difficulties, however, made the interpretation of these data somewhat dubious. Early in the course of the experiment it became obvious that a majority of the obtained deflections were going to be relatively small (10 or less). It likewise became evident that the (apparently) unavoidable inertia in the pointer apparatus was preventing the smaller deflections from being recorded at all, and was materially reducing the apparent magnitudes of the other deflections. Hence the small percentage of measurable deflections is partly spurious, as are also the small means.

Even more important in some respects than these failures of the recording mechanism to function correctly was another factor, the existence (but not the seriousness) of which was recognized from the very outset. Quoting from Wechsler's discussion of this factor (24):

In addition to the variations in the form of galvanic response [i.e., normal, multi-phasic, etc.] there is one general phenomenon which deserves special mention. It is the so-called "exhaustion" phenomenon, or diminution of the intensity of the galvanic response following its repeated or prolonged provocation. Thus the intensity of the responses which a subject gives at the beginning of an experiment are on the whole very much greater than those which he gives at the end, even though the emotions experienced may be more intense. Again, if an experiment is continued for a sufficient length of time, the response after becoming smaller and smaller will be unobtainable al-

together after a certain point. This phenomenon is observed most easily when the same stimulus is employed. With the repetition of a stimulus, its effectiveness in provoking a response tends to diminish progressively until after a little time it is entirely incapable of provoking a response. The psychogalvanic reflex thus seems, so to speak, to exhaust itself.⁶ (Pages 35-36.)

The term "positional effect" will be used to refer to this characteristic tendency of the deflections to decrease in magnitude with successive responses to similar stimuli. Sears (19) has called much the same phenomenon by the name "curves of adjustment", and has noted its occurrence after certain "critical points" where the intensity of the stimulus changes. In the present experiment, the opposite orders of the items in belts 1 and 4 was adopted in the attempt to compensate for the expected positional effect.

Figure 9, on the following page, shows how the positional effect operated in the first experiment. Figure 9 is what will be called a "positional" curve. The ordinates are obtained by averaging all of the measures (in this case, deflections) which are taken at corresponding positions in the series of stimuli (represented by the abscissae). Since at least one personality item and one achievement item are represented at each position, differences due to the type of material tend to cancel each other, and the resulting curve therefore becomes an index of the positional effect only.

Perhaps the value of this sort of curve should be explained more extensively. In the above quotation from Wechsler, it was stated that the exhaustion phenomenon "is observed most easily when the same stimulus is employed." Essentially the same statement also applies to successive repetitions of discrete but very similar stimuli. The stimuli might be considered simply as brief questions, disregarding for the moment the division into personality and achievement types. If the effects of the responses to these questions are attributable only to the fact that they are questions, and if these effects are not related in any way to the qualitative division of the questions into personality and achievement types, then the curves obtained by averaging the measures for the two types taken separately should not differ markedly from each other or from the positional curve representing the averages of all measures for both types. On the other hand, if the qualitative

⁵I.e., deflections accompanying the responses to personality items. The terms "personality" and "achievement" will be used in this adjectival manner, thus avoiding a profusion of awkward expressions like the foregoing.

⁶Wechsler, D. "The Measurement of Emotional Reactions", *Archives of Psychology*, XII, No. 76, pp. 35-36. Quoted by permission of the Editor.

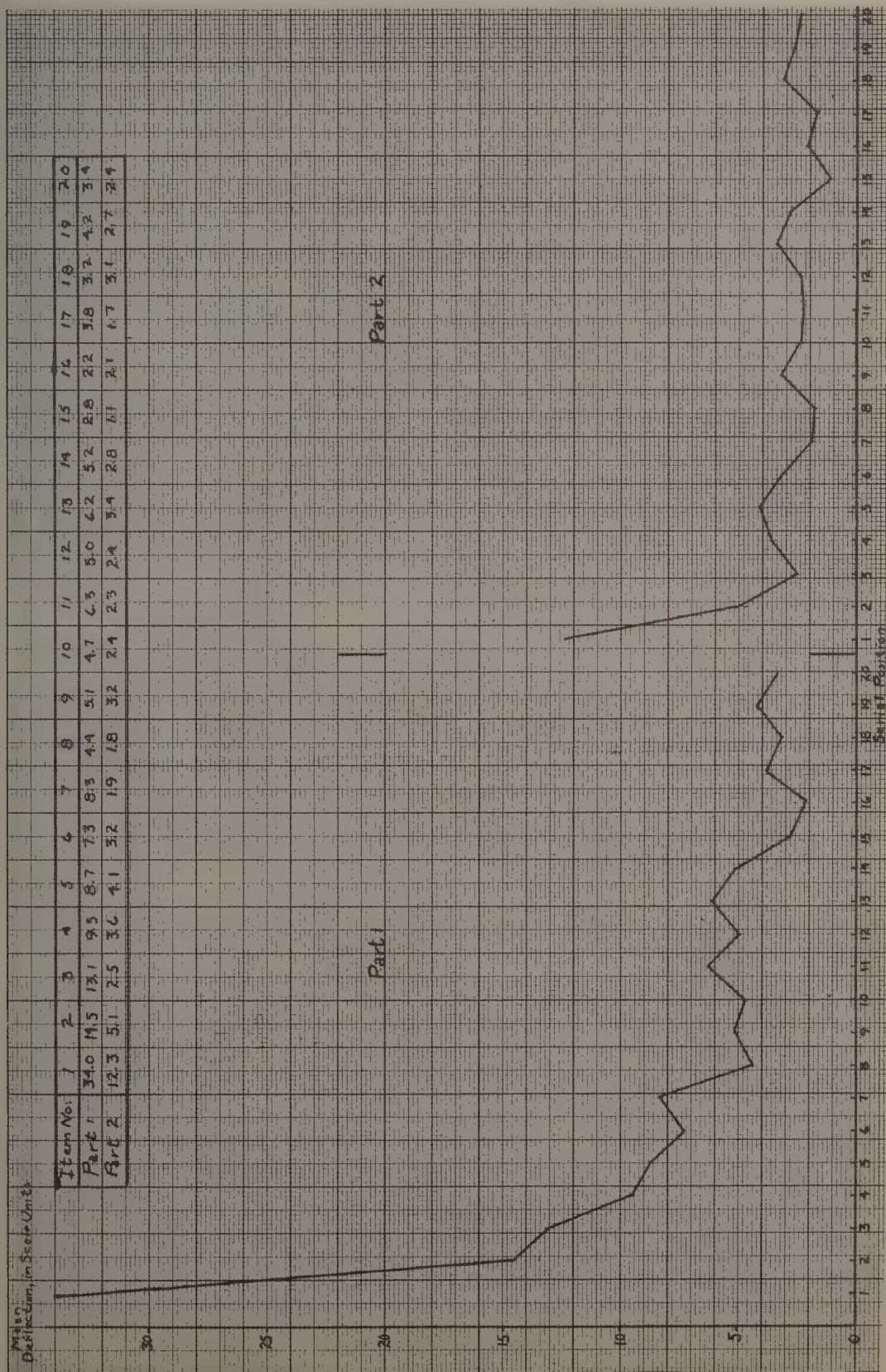


Figure 9. Positional Deflection Curves, Experiment 1. (Belts 1, 4, and 5 combined.)

differences between the two types are actually accompanied by different effects on the given measures (deflections in this case, although a limited use will be made of the positional curve in the treatment of reaction times and memory values in experiment 2), then the separate curves should be differentiated sharply from each other and from the positional curve.

Figure 9 demonstrates plainly that the positional effect was operating strongly in the first part of the series, and was repeated somewhat less strongly in the second part of the series. In both parts there is an initial high point, a rapid fall, and finally an approximate level. It appeared therefore that the rest period between parts 1 and 2 was equivalent to beginning all over again on a reduced scale.⁷

In spite of the small number of subjects represented in the foregoing curve, the suspicion grew that nothing was being demonstrated except the positional effect, or at least nothing which could be attributed to the different types of items. Two explanations were recognized as possible solutions of the dilemma: either (1) there were no differences, or (2) the way in which the stimuli were presented made it impossible for the differences to show up.

It seemed more hopeful to act on the assumption that the second of the foregoing possibilities was the one corresponding to the actual facts of the case. The immediate problem then became one of rearranging the conditions so that they would be favorable for the appearance of any differences which might actually exist.

One of the first steps, obviously, was to make sure that all of the valid deflections which occurred were actually recorded in their true magnitudes. It was for this reason that the pointer-tracing system was abandoned. As a matter of fact, the pointer was not used in taking the last five records included in the results of experiment 1, but these records formed too small a portion of the total to affect the averages noticeably.

Secondly, the positional curves of Figure 9 suggested that if any large differences were to show up, they would most likely be manifested on the relatively horizontal portions of the curves following the apparent exhaustion of the worst of the positional effect. Moreover, it would be highly desirable to concentrate the positional effect, insofar

as possible, in the first few positions of the series, rather than having it repeated in the middle. This meant the abolition of the rest period between Part 1 and Part 2.

The first attempt at rearrangement of the order of presentation of the items (called belt 1n) consisted of little more than the continuous presentation of the items as in belt 5, with no break in the middle and a single recall period at the end. Only 3 subjects were tried with the belt, but there seemed to be no improvement, and it was abandoned without further consideration.

Meanwhile, a more fundamental revision was being worked out. The curves of Figure 9 suggested that the major portion of the positional effect was concentrated within not more than 10 positions at the beginning of each series. Also, the positional effect was repeated when the type of item changed at the beginning of Part 2; and since this might conceivably have something to do with the nature of the item as well as the fact of change, it was deemed desirable to exhaust the positional effect as far as possible for both types of items before setting the stage for showing up any differences which might exist. These two ideas were made concrete in the arrangement of the items in Parts 1 and 2 of belts 2n and 3n, namely, 10 items of one type followed by 10 items of the other type.

Consideration of results such as those presented by Sears (19) further suggested that the differences, if at all real, would show to best advantage if a number of "critical points" were established in the later part of the series after (presumably) the greater share of the positional effect had been exhausted. With this thought in mind, Part 3 in belts 2n and 3n was composed of items of both types alternated irregularly, as was the whole of belt 5. It was hoped that, in spite of the failure of belt 5 to show anything significant, the changes from one type to the other might be more effective if they were withheld until the positional effect had been eliminated or largely exhausted. The same general principles also suggested the inclusion of an unexpected shift from a succession of items of one type to a succession of items of the other type, and this sort of shift was made the feature of Part 4 in belts 2n and 3n.

Reaction Times in Experiment 1. Whenever the reliabilities of obtained differences are to be found in this study, the computations will be based on the average results for

⁷The mean deflections for the separate belts were: belt 1, 4.07; belt 4, 7.70; and belt 5, 3.37. These differences make comparison of the separate item curves impossible, and are mentioned here in contrast with the corresponding situation in the second experiment, where the average deflections were very nearly equal for the two belts used.

individual subjects. Garrett (8, page 142) states that reliabilities should not be calculated when N is less than 25. The adoption of this standard precludes the calculation of reliabilities from the data of experiment 1. Therefore the reaction time results will be stated only in general terms, also the memory value results.

The 20 subjects of experiment 1 gave 800 responses, of which 788 resulted in usable reaction time records. The grand average was 6.56 seconds. The average personality reaction time was 6.48 seconds, and the average achievement reaction time was 6.62 seconds. As far as they go, these results agree with the difference in times for the group tests. However, if these average reaction times are multiplied by 125 (the number of items in each of the group tests) and change the seconds to minutes, the result is 13.79 minutes for the achievement test and 13.50 minutes for the personality test, a difference of .29 minutes, which is only a fraction of the actual difference (2.28 minutes) between the group test times.

Memory Values in Experiment 1. As in the case of the reaction times, no attempt will be made to consider the reliabilities of differences in this section.

The 20 subjects of experiment 1 recalled a total of 280 items, of which 153

were personality and 127 were achievement. For the personality items the data are: mean, 7.65; S.D., 1.78; V , 23.2. For the achievement items the data are mean, 6.35; S.D., 1.95; V , 30.7.

The individual subject performances agree very well with the group averages, since 14 of the 20 subjects recalled more personality items, 3 the same number of both types, and only 3 recalled fewer personality items than achievement items. For those who recalled more personality items, the average excess was 2.21; while for those who recalled more achievement items, the average excess was 1.66.

Significance of the Experiment 1 Results. At this point the findings of experiment 2 may be anticipated, by noting that there is general agreement between the two experiments with regard to the findings just presented; i.e., larger mean deflections, shorter reaction times, and higher memory values characterize the responses to the personality items in comparison with the responses to the achievement items. Since this agreement exists, and since no reliabilities have been calculated in experiment 1, all interpretations will be deferred until the corresponding results of experiment 2 have been presented.

Chapter IV

PSYCHOGALVANIC RESPONSES IN EXPERIMENT 2

Subjects. The findings presented in this chapter are based entirely on the records of 50 subjects who took the individual test after the changes had been made in the order of presentation of the items. There were 30 women and 20 men. The sexes were represented equally in belts 2n and 3n (10 men and 15 women in each). They were also represented equally in the vocal and non-vocal groups. This equality of distribution is important, since the data indicate that there are probably definite differences due to sex and to vocalization of answers.

General Results. The 50 subjects gave 1995 usable responses, of which 1663, or 83.1%¹ were accompanied by deflections large enough to be recorded (1 or more scale divisions). The difference between the 83.1% recorded in experiment 2 and the 50.6% recorded in experiment 1 is probably due in part to the new arrangement, but it reflects chiefly the greater efficiency of the direct recording method in comparison with the pointer apparatus.

The personality items were accompanied by measurable deflections in 838, or 83.8% of the cases, and the achievement items in 825, or 82.5% of the cases. This difference agrees with the corresponding difference found in experiment 1, but is too small to carry much significance. Figure 10 shows the distribution of all deflections measured in experiment 2. The zero deflections are indicated separately on the vertical axis, although properly a part of the lowest (0-4) interval. If 14 on the scale is taken as an arbitrary upper limit for "small" deflections, it is found that 78.2% of the personality deflections and 82.4% of the achievement deflections lie below this limit. Conversely, 21.8% of the personality deflections and 17.6% of the achievement deflections lie above this limit. Hence there is some tendency for the smaller deflections to be found with greater frequency in the achievement column, and vice versa. Two thirds of the very large deflections (60 scale divisions or more) are of the personality classification.

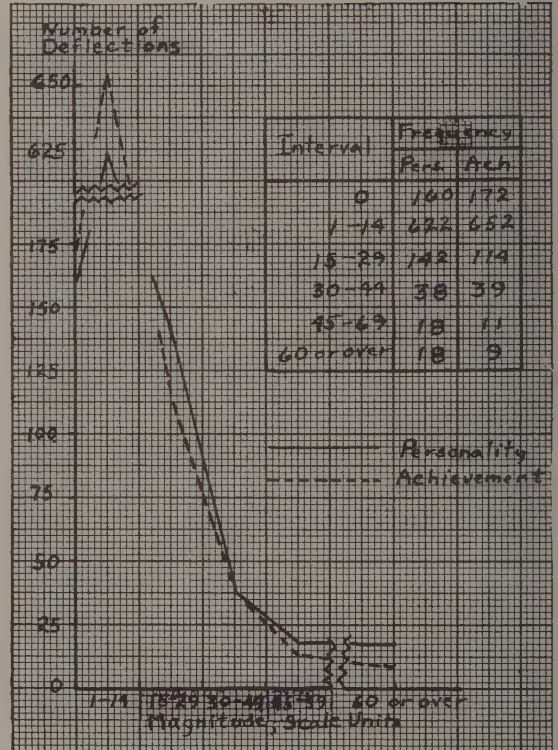


Figure 10. Distribution of All Deflections.
Experiment 2

Individual Subject Average Deflections. The reliability of the measures will be somewhat increased if instead of dealing with the original deflections the composite measures obtained by averaging for each subject the deflections falling into any classifications which we wish to consider are used. Figure 11 shows the distribution of the average personality deflections and the average achievement deflections (mean in each case) given by the 50 subjects. As was the case in Figure 10, the personality averages are slightly more numerous above the 14 limit, while the achievement averages are slightly more numerous below that limit, and vice versa.

¹Calculated on the basis of 2000 possible responses.

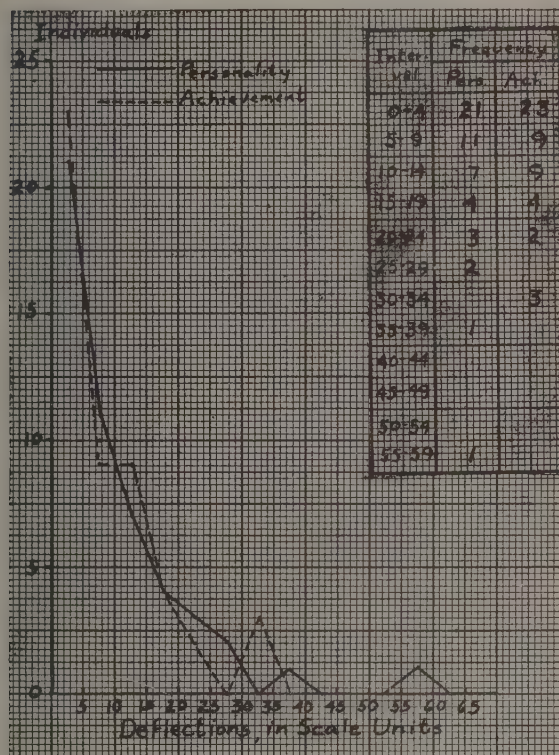


Figure 11. Distribution of Individual Subject Mean Deflections. Experiment 2

It might be noted here, parenthetically, that any differences which may show up in these individual averages are especially important, inasmuch as the positional effect definitely favored each type of item in half of the cases. Belt 2n began with 10 achievement items, thus giving this type the advantage of the relatively large initial deflections which almost invariably occurred. Belt 3n began with 10 personality items, thus giving that type the advantage. If the advantage gained from the initial positions in the series operated with equal force regardless of the type of item, it might be expected that all belt 2n subjects would have larger average achievement deflections, and that all belt 3n subjects would have larger average personality deflections. As a matter of fact, such a sharp division did not occur. Eight of the belt 2n subjects had larger average personality deflections in spite of the advantage of initial position held by the achievement items, while only three belt 3n subjects had larger average achievement deflections. Hence 30 (or 60%) of the subjects had larger average personality deflections,

19 (or 38%) had larger average achievement deflections, and 1 (or 2%) had equal average deflections for both types of items.

Individual Item Average Deflections.

The general tendencies of the personality and achievement deflections may be investigated from still another angle by considering the average of the deflections given by all of the subjects to each of the 20 personality items and each of the 20 achievement items. The effect of serial position is minimized in these averages, since each item appeared in the first half of one belt and in the last half of the other belt. Figure 12 shows the distribution of these individual item average deflections. The general tendency agrees with that found in Figures 10 and 11, since the personality item averages are more numerous in the higher intervals, while the achievement item averages are more numerous in the lower intervals. More precisely, 14 (or 70%) of the achievement items have average deflections smaller than 9 scale divisions, while only 7 (or 35%) of the personality items have average deflections falling within the same range. In the higher intervals the situation is reversed, 65% of the personality item averages being found there and 30% of the achievement item averages.

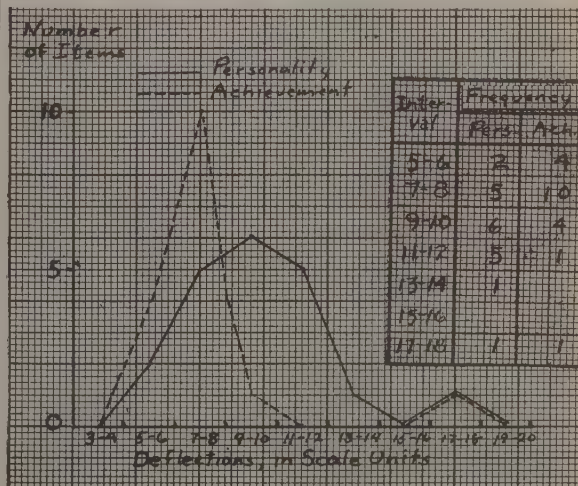


Figure 12. Distribution of Individual Item Mean Deflections. Experiment 2

The Positional Deflection Curve, Experiment 2. It will be remembered that the changed order of items used in the second experiment had as one of its purposes the elimination of the positional effect from at least a part of the series, by attempting to confine it within Parts 1 and 2. It is pertinent at this point to inquire into the evidence

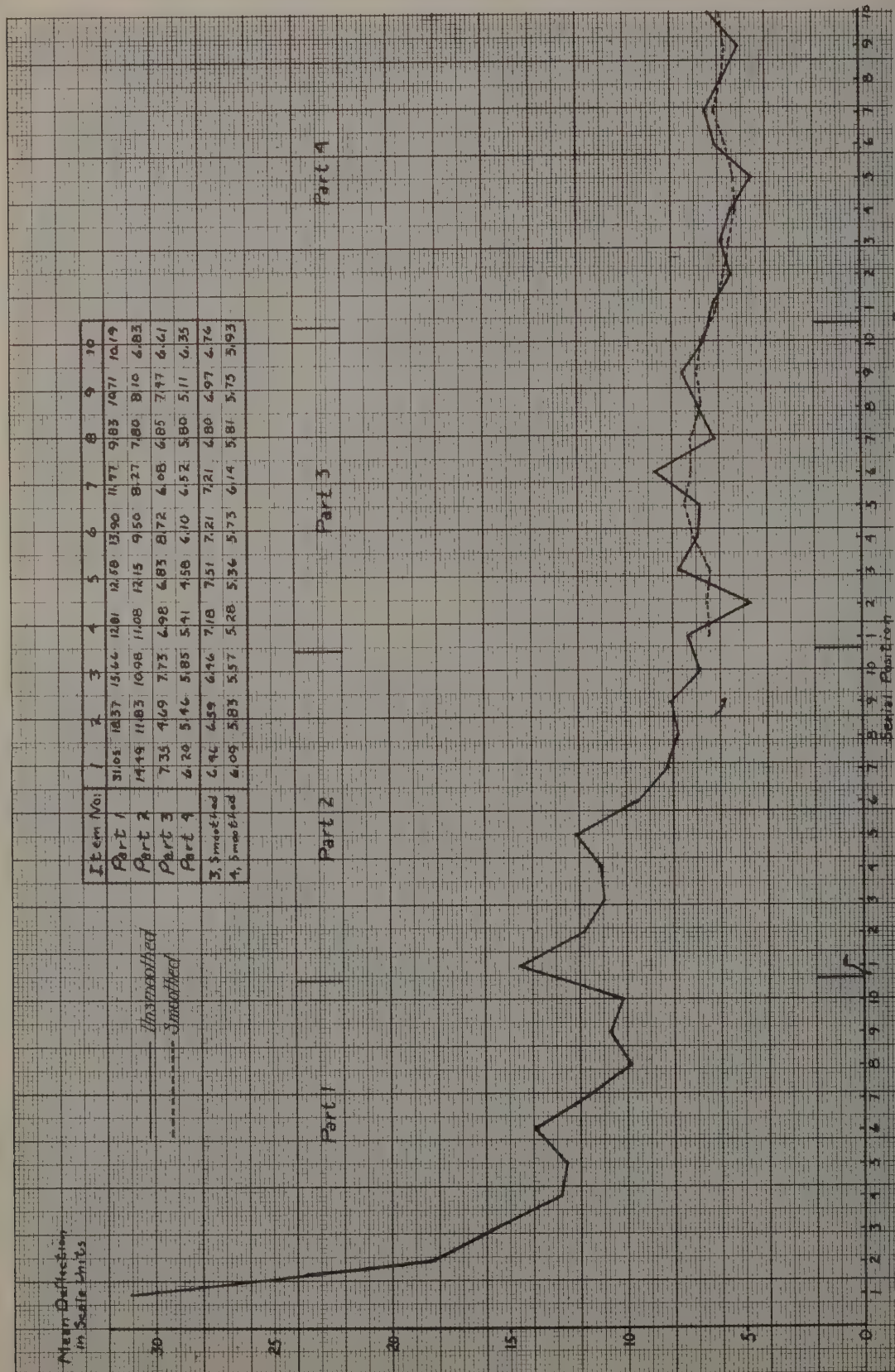


Figure 13. Positional Deflection Curve, All Subjects. Experiment 2

concerning the degree of success which this attempt met.

Figure 13 shows the positional curve for the deflections measured in experiment 2. There is a rapid fall in the curve at the beginning of Part 1, a less rapid fall after the third position, and a rise at the beginning of Part 2 followed by a slow fall for the remainder of Part 2; but in Parts 3 and 4 the curve is relatively horizontal, at least in comparison with Parts 1 and 2. In order to show this latter tendency more clearly, the curve through Parts 3 and 4 has been smoothed by the three-point-average method (Rugg, 18, page 184).

It is evident that, for the group of subjects as a unit, the major portion of the positional effect actually has been confined to the first half of the stimulus series, otherwise the flattening out would not occur. Therefore we may assume that any differences between the average deflections for the two types of items are affected only to a slight degree by the positional factor, when such differences are calculated from measurements taken in the last half of the series. On the other hand, differences manifested in the first half of the series are bound to be com-

plicated by the positional effect, although it would be reasonable to expect them to show the same general tendencies found in the more favored second half of the series.

Personality and Achievement Deflection Curves in Parts 3 and 4. The deflection curves for the separate types of items, and for the two belts, will now be compared with each other in Parts 3 and 4 where the positional effect may be disregarded.

Figure 14 shows the curves for the two belt groups, without the positional curve. The values (mean deflections, of course) for the first position of Part 3 are nearly equal. From that point on, however, the curves behave in a most interesting fashion. Both curves fall at the second position, but the a-p (i.e., achievement item to personality item) fall is less than the p-a fall. Beginning at this point in the series, there are 6 p-a shifts, 6 a-p shifts, 1 a-a shift, and 1 p-p shift. Inspection of the curves shows that 5 of the 6 p-a shifts result in smaller achievement values, and also 5 of the 6 a-p shifts result in higher personality values. The two values involved in the p-p shift are both relatively high, and the two values involved in the a-a shift are both relatively

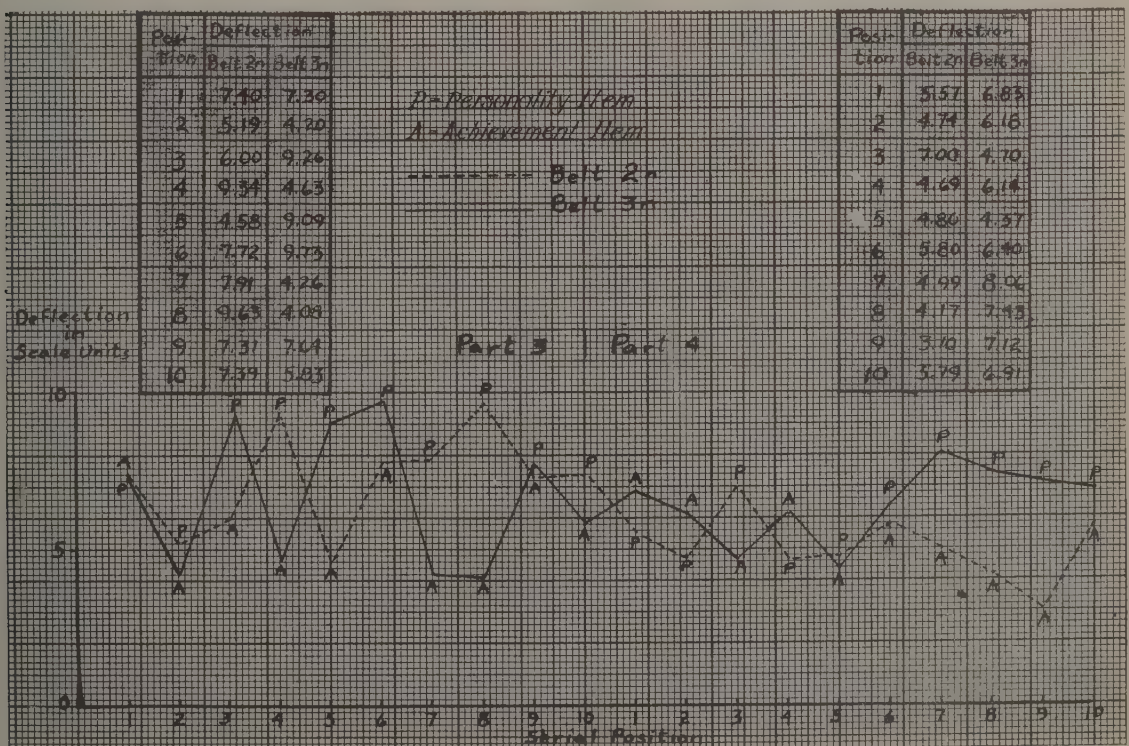


Figure 14. Mean Deflection Curves, Belts 2n and 3n. Experiment 2

low. It is thus seen that there was a tendency (perfect in belt 3n, strong in belt 2n) for p-a shifts to produce a falling off in the mean deflection and for a-p shifts to produce a rise in the mean deflection.

Another "critical" point came at position 6 of Part 4. During the first 5 positions of this Part, when the subject knew the type in advance, the two belt curves criss-cross irregularly (as in Parts 1 and 2, for which the curves have not been constructed); but beginning with the unexpected shift at position 6 the curves separate sharply, and for the remainder of the series the mean personality deflections are uniformly higher than the mean achievement deflections.

Figure 15 is derived from Figure 14 by connecting all of the personality values to form one curve, and all of the achievement values to form another curve. In Part 3 (except for the first position) all of the personality values are higher than both the achievement curve and the positional curve, and the same is true in the last half of Part 4. Figure 16 is the same as Figure 15, except that the general tendencies have been emphasized by smoothing the curves.

Measures of Central Tendency. The

data presented up to this point have indicated that the average personality deflections seem to be somewhat larger than the average achievement deflections. It now becomes necessary to express this difference more precisely and test it for reliability (that is, for freedom from the sampling error). The C.R. or critical ratio has already been used and defined. The formula for computing the standard deviation of the difference between two averages (used in computing the C.R.) is:

$$S.D.(diff.) =$$

$$\sqrt{S.D.(aver_1)^2 + S.D.(aver_2)^2 - 2r S.D.(aver_1)S.D.(aver_2)}$$

(Garrett, 8, p. 286f.)

Where the measures are not correlated, the third term under the radical drops out.

The individual subject averages will be used in all reliability calculations rather than the original ungrouped measures, and means rather than medians. The highly skewed nature of the deflection distributions raises the question as to which measure of central tendency (mean or median) would be best in this case. The median is less affected by

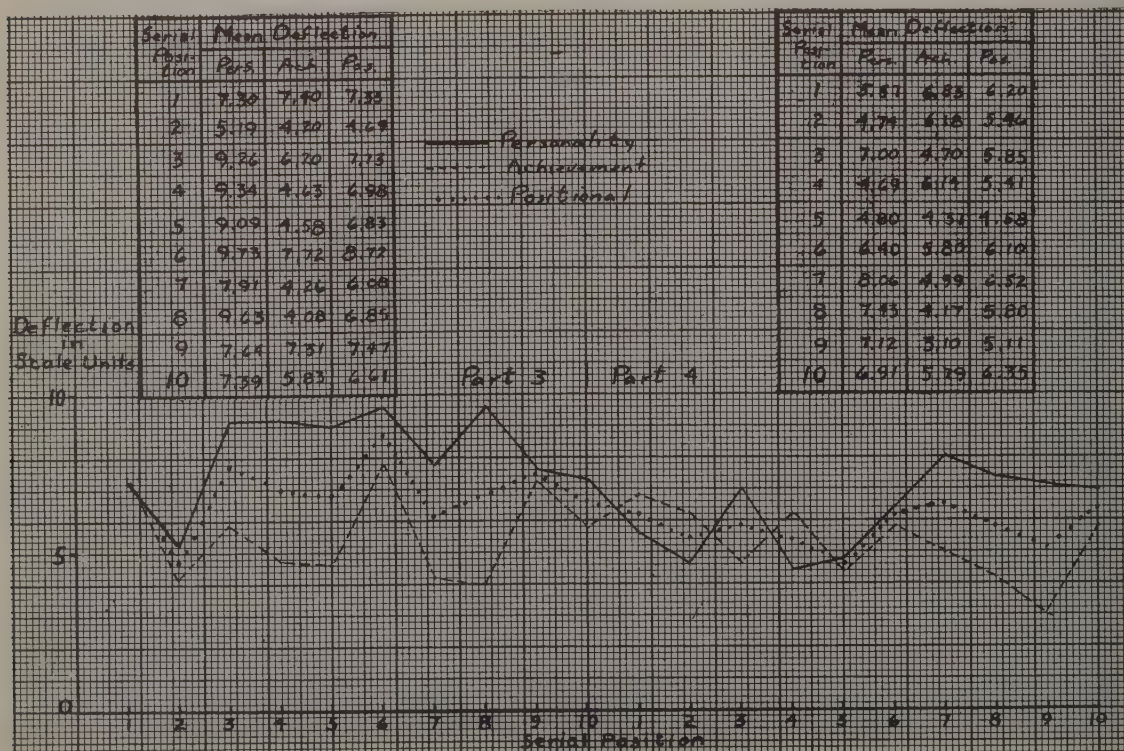


Figure 15. Personality, Achievement and Positional Deflection Curves, Parts 3 and 4. Experiment 2

large outlying scores than the mean. However, there is reason for believing that the difference, if any, between the personality and achievement deflections may be due precisely to the greater frequency of the very large personality deflections. Such being the case, the mean will be used as being the measure which is more sensitive to this sort of difference. The mean will be calculated directly (sum of the measures/N) rather than from frequency distribution.

However, if there are any real differences, the central tendencies ought to agree in direction, if not in magnitude, regardless of the method of computation. Table VI compares the central tendencies of the deflections in several of the major categories dealt with. The "true means" are the averages derived from the original ungrouped deflections. The "F.D. Means" are computed from the original deflections grouped in frequency distributions. The "Ind. Means" are the averages of the individual subject averages. The Medians were computed from the frequency distributions used for the F.D. Means.

Table VI shows that, with a single exception, the average personality deflection is larger than the average achievement deflec-

tion in each classification and by each of the four methods of computation. The sole exception is the larger median non-vocal achievement deflection, but this is also the smallest difference in the whole Table. The means calculated from frequency distribution are regularly the largest, and the medians the smallest of the averages. The individual subject means agree most closely with the true means, and will be used in the reliability calculations. This choice gives a convenient number (20 to 50) cases to work with, it facilitates the computation of correlations where they exist, and it makes the psychogalvanic results comparable with the reaction time and recall measures which are also calculated on the individual subject basis.

Reliabilities of Differences. The Tables, Figures and discussions of the remainder of this chapter present the results obtained by comparing the individual subject average personality deflections with the individual subject average achievement deflections, classified from a number of different angles. Excepting where it is necessary to consider the men as a group, not fewer than 25 cases (Garrett's minimum) will be used in any reliability calculation. The general form of

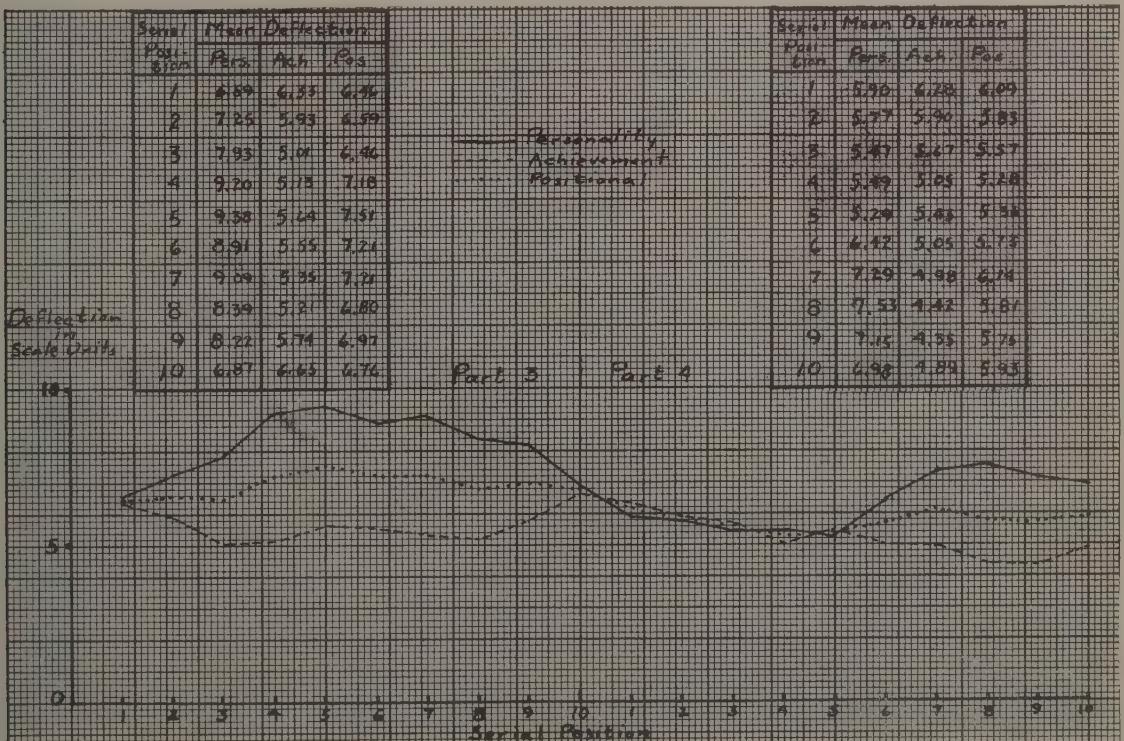


Figure 16. Smoothed Deflection Curves. (From Figure 15)

Table VI
AVERAGE DEFLECTIONS IN EXPERIMENT 2, CALCULATED IN
FOUR DIFFERENT WAYS

Classification	True Mean	F.D. Mean	Ind. Mean	Median
All deflections	9.33	10.40	9.34	5.33
All personality	9.90	10.72	9.89	5.46
All achievement	8.77	9.58	8.80	5.12
All vocal	12.65	15.21	12.72	8.46
All non-vocal	6.12	6.89	5.96	3.99
Vocal personality	13.53	16.30	13.51	8.66
Vocal achievement	11.77	12.55	11.80	7.99
Non-vocal personality	6.26	7.15	6.26	3.98
Non-vocal achievement	5.77	6.62	5.87	4.02
Part 1 personality	14.96	15.80	15.00	9.89
Part 1 achievement	14.39	14.88	14.40	9.24
Part 2 personality	10.20	11.13	10.20	6.08
Part 2 achievement	9.91	10.69	9.91	5.77
Part 3 personality	8.24	9.14	8.21	4.49
Part 3 achievement	5.59	6.56	5.60	4.39
Part 4 personality	6.21	7.28	6.14	4.25
Part 4 achievement	5.14	6.39	5.25	4.16

the Tables has already been described and explained.

Inter-type vs. Intra-type Differences.

It was stated in an earlier chapter that the personality-achievement differences might be significant in themselves (inter-type differences) or perhaps greater than differences between classes within the same type (intra-type differences). Table VI shows very clearly that the differences within the types (such as the difference between the vocal personality deflections and the non-vocal personality deflections) have a range which is considerably greater than the range of the personality-achievement differences. It will be shown later that sex and vocalization are apparently the factors responsible for this state of affairs.

The Personality-Achievement Difference for the Entire Stimulus Series. This comparison is based on the individual subject average deflections for all items.

Table VII
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE
BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS
FOR (1) ALL PERSONALITY ITEMS, AND (2) ALL
ACHIEVEMENT ITEMS
(50 Subjects)

Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff.	(diff)	r	C.R.
(1) 9.892	10.41	105.23	1.47	1.086	.80	.84	1.35
(2) 8.806	8.27	93.91	1.17				(91)

Complete reliability demands that the C.R. should be 3.00 or greater. The difference in Table VII has 91 out of 100 chances of being a real difference greater than zero. (Note that the 91 is placed in parentheses directly under the C.R. of which it is the interpretation.)

The very large S.D.'s indicate that there is a wide range of individual differences, and the V's further show that the range is relatively greater for the personality deflections than for the achievement deflections. Since 0 is the lower limit for both distributions, the greater variability of the personality deflections indicates a proportionately larger number of personality deflections in the higher intervals. The same line of evidence also makes it plain that a few of the more strongly reacting subjects are contributing more than their fair share to the group averages. The correlation of .84, however, shows that the same individuals who give large personality deflections also give large achievement deflections. Accordingly, the stronger reactivity of these subjects tends to inflate the averages but not to alter their directional relationships.

A basic issue at this juncture concerns the conditions (if any) under which it is valid to compare and combine the psychogalvanic responses given by different subjects. For the purposes of this investigation, the essential soundness of Wechsler's method has been assumed. That method has not gone unchallenged, but neither has any other. The present writer is not prepared to debate the question, since the psychogalvanic response itself is not the primary object being here investigated, but is being used in a purely instrumental fashion as one of several possible differentiating agents by means of which the subjective conditions influencing the responses to the two types of test items might be studied. The two types of items were cast in the same form with the same alternative answers. They were presented to each subject at a single sitting, by means of the same apparatus, and otherwise under as nearly the same external conditions as could be devised. It is therefore difficult to see how there could possibly be any factor in the experimental situation which would act consistently on one type of stimulus and not on the other. Consequently, it is believed that the subjective effects of the two types of items constituted the major variable affecting the subjects' responses, and that any obtained differences in psychogalvanic deflections or the other measures may rightly be attributed to these respective subjective effects. If there are any serious errors involved in the

combination of deflections given by different subjects, the same errors appear in both members of each comparison which is made, and should therefore tend to be neutralized.

The discussion now returns to the series of direct comparisons between various dichotomous classifications of personality and achievement deflections.

The Personality-Achievement Difference for Parts 1 and 2 Considered as a Unit. Table VII showed that for the series as a whole, the obtained difference was too small to be completely reliable, yet too large to be entirely without significance. If the previously stated suspicions regarding the probable influence of the positional effect on any differences which might exist between the personality and achievement deflections are correct, then the difference ought to be least reliable in the first half of the stimulus series, where the major portion of the positional effect is concentrated. The pertinent data are given in Table VIII.

Table VIII

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR

- (1) THE PERSONALITY ITEMS OF PARTS 1 AND 2, AND
(2) THE ACHIEVEMENT ITEMS OF PARTS 1 AND 2

(50 Subjects)

	Average	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	12.604	12.91	102.43	1.82	.446	1.23	.77 .36
(2)	12.158	12.92	106.26	1.82			(64)

The findings confirm the suspicions. The positional effect has brought the means very close together, and it has increased the variability of the achievement deflections to a value higher than that of the personality deflections (whereas for the whole series and for Parts 3 and 4, the achievement deflections are less variable). The C.R. shows that the obtained difference is not much better than chance.

The Personality-Achievement Difference for Part 3. In contrast with Parts 1 and 2, Part 3 might reasonably be expected to show the most reliable difference of any portion of the series, especially in view of the definite separation found between the personality and achievement curves of Figures 15 and 16. Table IX gives the data pertaining to the reliability of the personality-achievement difference for the deflections of Part 3.

Table IX

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR

- (1) THE PERSONALITY ITEMS OF PART 3, AND (2) THE ACHIEVEMENT ITEMS OF PART 3

(50 Subjects)

	Average	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	8.21	10.35	126.06	1.46	2.61	1.17	.61 2.23
(2)	5.60	4.86	86.78	.68			(99)

While still falling short of complete reliability, the Part 3 difference contrasts sharply with the differences for the whole series and for Parts 1 and 2. It is much larger both absolutely and relatively, and this fact, together with the sharp increase in the C.R., makes it much more probable that it is a real difference.

The Personality-Achievement Difference for Part 4. The curves of Figures 15 and 16 do not lead to the expectation of as large or as reliable a difference for Part 4 as was found for Part 3. The near coincidence of the two curves for the first five items of Part 4 represents what is probably a distinct relaxation from the relative tension created by the several successive critical points of Part 3; and even the sharp separation of the curves after item 5 cannot be expected to overcome the handicap presented by the practically equal central tendencies of the first 5 items. The data concerning the reliability of the personality-achievement difference for the Part 4 deflections are given in Table X.

Table X

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR

- (1) THE PERSONALITY ITEMS OF PART 4, AND (2) THE ACHIEVEMENT ITEMS OF PART 4

(50 Subjects)

	Average	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	6.14	7.86	128.01	1.11	.89	.87	.62 1.02
(2)	5.25	4.71	89.71	.66			(84)

The difference is small in comparison with the Part 3 difference, and the reliability, while higher than in Parts 1 and 2, is not as high as in Part 3 or the whole series.

The Personality-Achievement Difference for the Last Half of Part 4. The curves

of Figures 15 and 16 indicate quite a large difference, relatively, between the personality and achievement deflections for the last 5 items of Part 4. However, different subjects are represented in the two curves, hence there is no correlation, and the smaller number of subjects (25) which has to be used in computing the S.D. (aver) for each type precludes the finding of any very high reliability. Nevertheless, as seen in Table XI, even in spite of these handicaps the difference for the last half of the Part stands up as well as the difference for all of Part 4.

Individual Records vs. the Group Averages. The group averages which have been used uniformly show larger values for the personality deflections than for the achievement deflections. To what extent do the individual subjects agree with the tendency for the group

Table XI
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS FOR THE LAST HALF OF PART 4: (1) PERSONALITY ITEMS, (2) ACHIEVEMENT ITEMS
(25 Subjects in Each Group)

	Average	S.D. (dis)	V	S.D. (aver)	S.D. Diff.	(diff)	r	C.R.
(1)	7.00	9.99	142.71	1.99	2.22	2.17	--	1.02
(2)	4.78	4.37	91.42	.87				(84)

as a whole? For the entire series (Table VII) 30 of the subjects had larger mean personality deflections, 19 had larger mean achievement deflections, and in one case the two were equal. For Part 3, the corresponding numbers were 28, 19 and 3. Thus in both of these classifications where the highest reliabilities are found, a majority of the individual subjects agree with the group tendency. In Parts 1 and 2 (Table VIII) the situation is reversed, 24 subjects agreeing with the group tendency, 25 disagreeing, and one having equal deflections for both types. The same was true of Part 4, where the corresponding numbers were 21, 24 and 5. In both of the latter cases, however, the reliabilities were lower (decidedly so for Parts 1 and 2) than in the cases where a majority of the subjects agreed with the group tendency, and in neither case did the disagreeing subjects form an actual majority of the total number.

There is a further comparison bearing on this matter of the correspondence between individual and group results. If only chance fluctuations were involved, the individual subject differences in cases where the average personality deflection was smaller than

the average achievement deflection should be just as large, on the average, as the differences in cases where the average personality deflection was larger than the average achievement deflection. Stated somewhat differently, the average excess of the mean personality deflections over the mean achievement deflections for those subjects who agreed with the group tendency should be about equal to the average excess of the mean achievement deflections over the mean personality deflections for those subjects who disagreed with the group tendency. The data show that such was not the case. For the whole series, the mean p-a difference was 3.8, while the mean a-p difference was 3.3. For parts 1 and 2 the mean p-a difference was 5.9, and the mean a-p difference was 4.8. For Part 3 the corresponding values were 6.3 and 2.4, while for Part 4 they were 5.3 and 2.7. In other words, at no time did the individual subject mean achievement deflections exceed the individual subject mean personality deflections by amounts which were, on the average, greater than the amounts by which the individual mean personality deflections exceeded the individual mean achievement deflections. Accordingly, even when the factors in the situation prevented an individual subject from agreeing with the group tendency, the achievement deflections were under a handicap which prevented them, on the average, from surpassing the personality deflections by amounts as great as the amounts by which the personality deflections surpassed the achievement deflections in the cases of those subjects who agreed with the group tendency.

Relative Variabilities of Personality and Achievement Deflections. The Coefficients of Variation given in Tables VII to XI have already been mentioned once or twice. Here attention is directed to them somewhat more specifically. Excepting in Parts 1 and 2, where the positional effect confuses all other tendencies, each of the personality V's is larger than the corresponding achievement V, and the range of the personality V's does not overlap with the range of the achievement V's. For all of the Tables, the average personality V is 120.9, and the average achievement V is 93.6; that is, the achievement deflections are approximately 77% as variable as the personality deflections. The responses to the personality items, therefore, seem to excite sympathetic activity over a wider range of intensities than the responses to the achievement items, indicating that the subjective factors influencing the psychogalvanic responses are likewise more variable under the stimulation furnished by the personality items. This is an important

difference in itself, regardless of the differences in the averages.

Average Deflections for the Individual Items. (Data corresponding with Figure 12.) The average of the 20 mean deflections for the separate personality items was 9.938, and the corresponding figure for the achievement items was 8.753, agreeing closely with both the true means and the individual subject means for the personality and achievement deflections. The respective variabilities for the personality and achievement item deflections are 29.2 and 27.4, indicating slightly greater variability for the personality item deflections. Since there are only 20 items in each group, no reliability data for the difference between the averages are presented.

Effects of Sex and Vocalization of Answers. Earlier in this chapter it was stated that both sex and vocalization appeared to be factors influencing the magnitude of the psychogalvanic responses to the stimuli used in this study. The evidence in support of these statements will now be considered, and also the effects of the sex and vocalization factors on the personality-achievement difference which is the chief object of interest.

The mean deflections for the various classifications wherein these sex and vocalization differences are to be studied are given in Table XII.

Table XII

INDIVIDUAL SUBJECT MEAN DEFLECTION AVERAGES FOR MEN AND WOMEN, AND FOR THE VOCAL AND NON-VOCAL GROUPS

Classification	Men	Women	Both
All deflections	12.591	7.175	9.341
All personality deflections	13.243	7.657	9.892
All achievement deflections	11.945	6.714	8.806
All vocal deflections	19.378	8.293	12.727
All non-vocal deflections	5.805	6.056	5.966
Vocal personality deflections	20.775	8.676	13.516
Vocal achievement deflections	17.990	7.682	11.808
Non-vocal personality deflections	5.712	6.638	6.267
Non-vocal achievement deflections	5.901	5.744	5.872

The first three figures in each column represent the average results for 20 men, 30 women, 50 in all. The six remaining figures in each column represent the results for 10 men, 15 women, 25 in all.

With a single exception (non-vocal men) Table XII confirms the personality-achievement difference found in the data of

Tables VII to XI. Two additional tendencies are even more apparent: (1) the average deflections given by the men tend to exceed the average deflections given by the women, and (2) the average deflections given by the vocal group tend to exceed the average deflections given by the non-vocal group.

One of the averages (non-vocal men, personality) is too small, and reverses both the sex and vocalization differences for that group and also causes a discrepancy in the sex difference for the non-vocal group as a whole. However, these inconsistencies involve only 10 men and (or) 15 women. All averages based on the deflections for all of the men and (or) all of the women agree with the general sex and vocalization tendencies noted above.

The Tables and discussions which follow will consider the reliabilities of the sex and vocalization differences in those cases where the results for all of the subjects can be included in the comparison.

The Sex Difference for all Subjects.

In making sex comparisons it is necessary to violate the requirement that N should be at least 25, since there were only 20 men.

Table XIII

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY (1) ALL MEN, AND (2) ALL WOMEN
(20 Men, 30 Women)

	Average	S.D. (dis)	V	S.D. (aver)	Diff. (diff)	r	C.R.
(1) 12.591	10.89	86.48	2.43	5.416	2.70	—	2.00
(2) 7.175	6.51	90.73	1.18				(98)

Figure 17 shows the distributions of the individual subject average deflections for the two sex groups. Since the numbers of men and women are unequal, the ordinates have been made comparable by expressing them as percents. It may be noted that nearly twice as many women as men are found in the lowest interval, that the men exceed the women in all of the higher intervals except one, and that 15% of the averages for the men are entirely above the range for the women. In terms of overlapping, the data show that 65% of the men have mean deflections larger than the average mean deflection for the women, while only 20% of the women exceed the average for the men. The C.R. is smaller than might be expected from the size of the difference, due to the smaller number of men, but the supplementary evidence from Figure 17 makes it extremely probable that there is

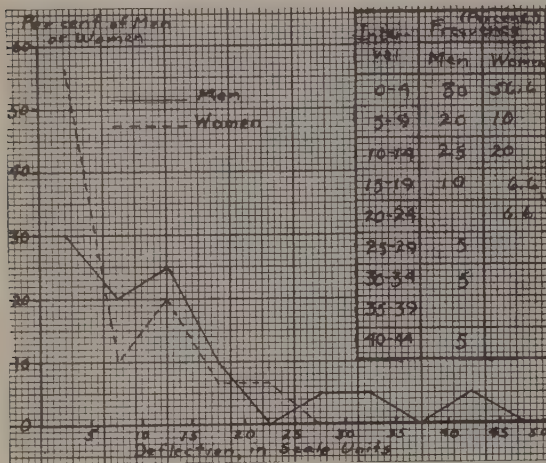


Figure 17. Distribution of Individual Mean Deflections for Men and Women. Experiment 2

here a real difference greater than zero. It may therefore be concluded that, under the conditions of this experiment, the average deflection given by the men is significantly larger than the average deflection given by the women.

The Vocalization Difference for All Subjects. The effect of vocalization of answers as compared with the silent method of responding will now be considered. Figure 18 shows the distributions of the original ungrouped deflections given by the vocal and non-vocal groups (comparable with Figure 10). The zero deflections are nearly two and one half times as numerous in the non-vocal distribution as in the vocal distribution, while every interval beginning with the one whose lower limit is 5 has a large excess of vocal deflections over non-vocal deflections, and the vocal deflections extend for several intervals higher than the upper limit of the range of the non-vocal deflections.

Figure 19 presents the same data in terms of the individual subject averages. The number of non-vocal averages found in the lowest interval is nearly twice the number of vocal averages found there, and 20% of the vocal averages lie entirely above the range of the non-vocal averages. In terms of overlapping, 17 (or 68%) of the vocal individual mean deflections are above the average of the non-vocal, while only 3 (or 12%) of the non-vocal individual mean deflections are above the average of the vocal.

Table XIV gives the reliability data for the obvious difference revealed by Figures 18 and 19, that is, much larger average deflections when the answers are given vocally.

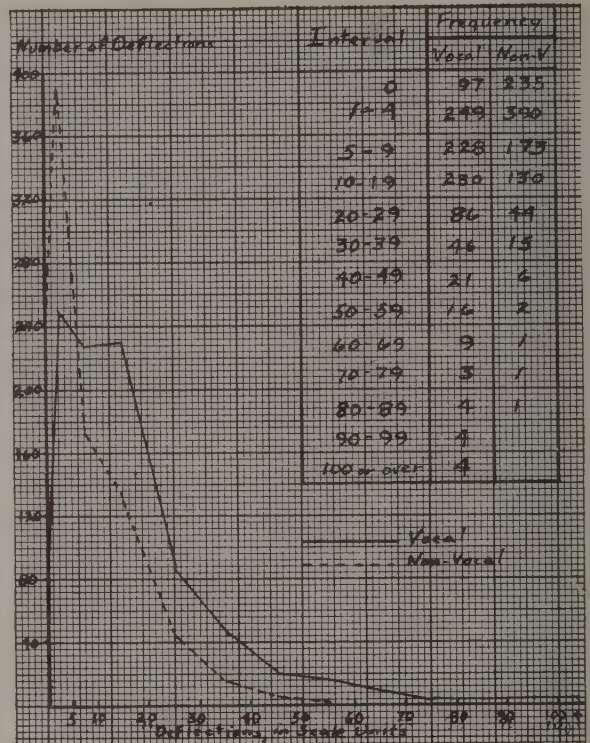


Figure 18. Distribution of Deflections Given By the Vocal and Non-vocal Groups. Experiment 2

Table XIV
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP (25 Subjects in Each Group)

	Average	S.D. (dis)	S.D. V	S.D. (aver) Diff.	(diff) r	C.R.
(1)	12.727	10.67	83.83	2.13	6.761	2.36
(2)	5.966	5.17	86.65	1.03		(99)

In terms of the C.R., the difference between the vocal and non-vocal mean deflections is the most reliable difference that has been found between any two classes of deflections. Further evidence concerning it is given in Figure 20 which shows the positional curves for the vocal and non-vocal groups contrasted with each other and with the general positional curve of Figure 13. The three curves never cross or meet, and they are essentially parallel, indicating that the vocalization difference is consistent throughout the entire stimulus series.

In spite, therefore, of the failure of the difference in the means to show

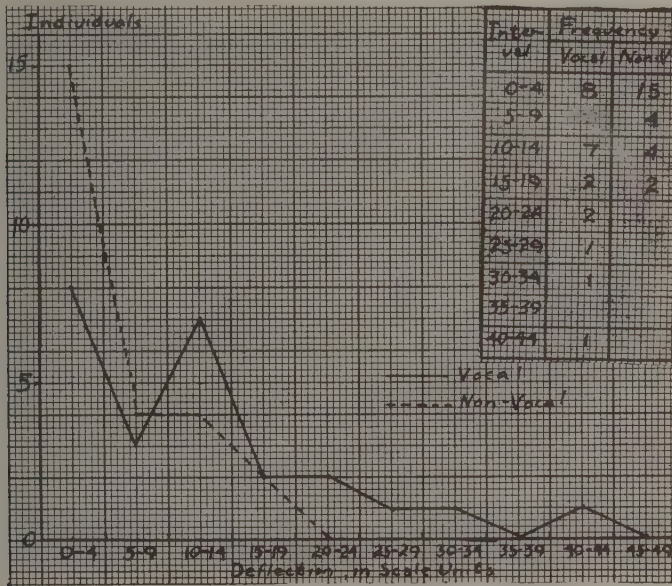


Figure 19. Distribution of Individual Mean Deflections for the Vocal and Non-vocal Groups. Experiment 2

complete reliability, it seems reasonable to conclude that the tendency for the deflections given by those who vocalized their answers to be larger than the deflections given by those who did not vocalize their answers represents a real difference.

Relationships Between the Sex Difference and the Personality-Achievement Difference. In connection with both the sex and vocalization differences, interest attaches to their effects on the personality-achievement difference which is the primary object of concern in this investigation.

Two issues are raised here: (1) is the sex-difference relatively larger or more reliable for either type of item, and (2) is the personality-achievement difference relatively larger or more reliable for either sex? Also, are there any differences in the relative variabilities?

Tables XV (a) and XV (b) present the data concerned with the first of these issues.

Table XV (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO PERSONALITY ITEMS BY (1) ALL MEN, AND (2) ALL WOMEN (20 Men, 30 Women)

	Average	S.D. (dis)	V	S.D. (aver)	Diff. (diff)	r	C.R.
(1)	13.243	13.39	101.11	2.99	5.586	3.25	1.71
(2)	7.657	7.02	91.64	1.28			(96)

²i.e., 96 and 99 chances out of 100 respectively.

The differences are relatively equal, each being about six tenths of the average deflection for the whole type. The reliabilities also are nearly equal.² The only difference of consequence seems to be that the personality deflections are more variable for the men, while the achievement deflections are more variable for the women.

It may be concluded, therefore, that the sex difference is practically independent (in magnitude, at least) of the type of item.

The second of the two issues is dealt with in Tables XVI (a) and XVI (b).

For the men, the personality-achievement difference is about 10% of the average deflection given by men to all items. For the women, the personality-achievement difference is about 13% of the average deflection given by women to all items. These facts, plus the larger critical ratio for the women, point superficially to a slightly higher reliability for the

Table XV (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO ACHIEVEMENT ITEMS BY (1) ALL MEN, AND (2) ALL WOMEN, EXPERIMENT 2 (20 Men, 30 Women)

	Average	S.D. (dis)	V	S.D. (aver)	Diff. (diff)	r	C.R.
(1)	11.943	9.05	75.76	2.02	5.231	2.37	2.20
(2)	6.714	7.05	105.00	1.29			(99)

Table XVI (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE MEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS (30 Subjects)

	Average	S.D. (dis)	V	S.D. (aver)	Diff. (diff)	r	C.R.
(1)	13.243	13.39	101.11	2.99	1.298	1.42	.91
(2)	11.945	9.05	75.76	2.02			(82)

personality-achievement difference with the women. The smaller number of men, however, renders this point uncertain. The variabilities are the same as in the preceding Tables.

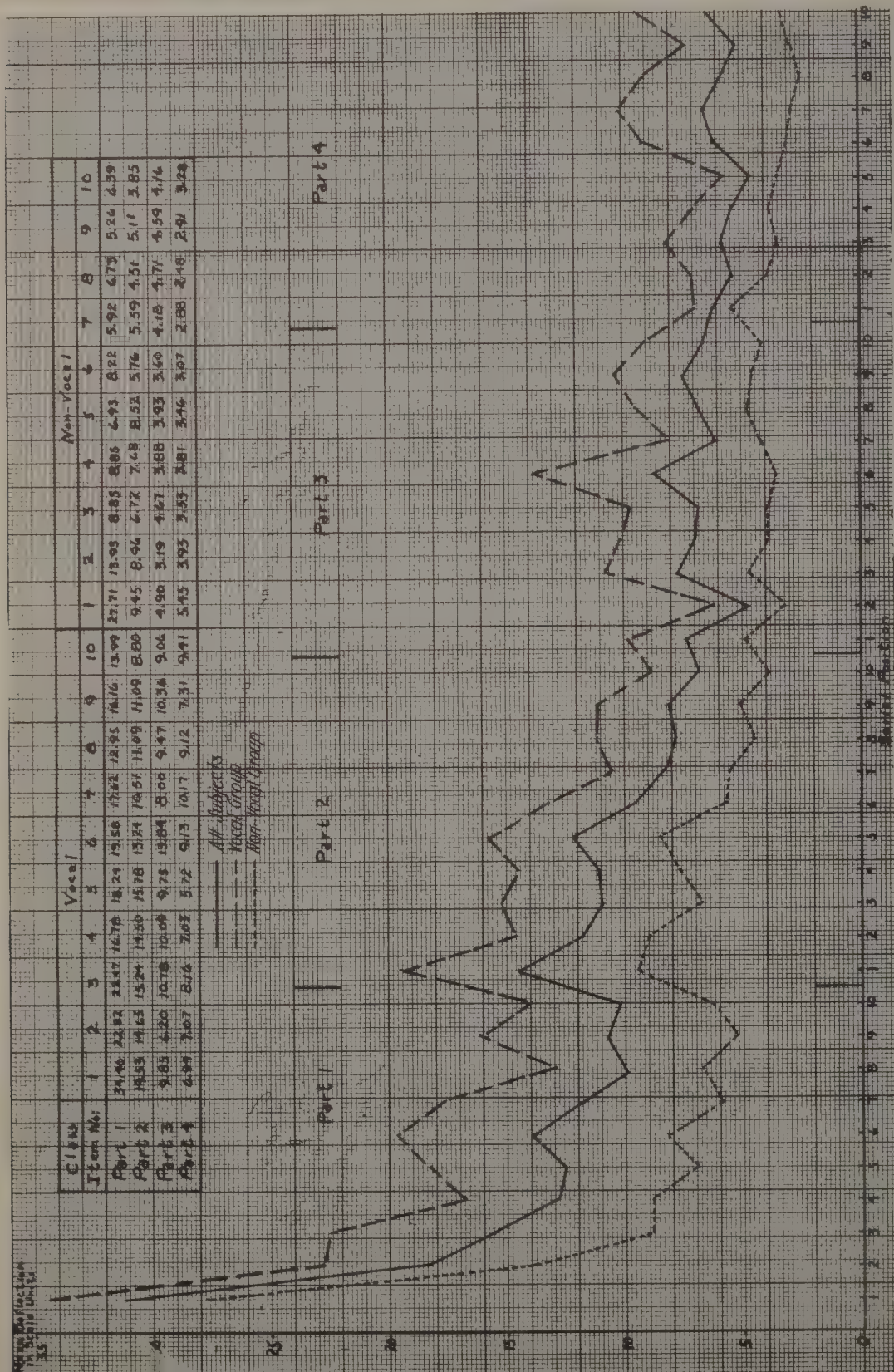


Table XVI (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE WOMEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS
(30 Subjects)

	Aver- age	S.D. (dis)	V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	7.657	7.02	91.64	1.28	.943	.726	.84
(2)	6.714	7.05	105.00	1.29			1.29 (90)

On the whole, it cannot be said that the personality-achievement difference is either markedly larger or more reliable for either sex, although it is perhaps very slightly so for the women.

Relationships Between the Vocalization Difference and the Personality-Achievement Difference. The issues to be raised here correspond with those in the preceding section, namely, (1) is the vocalization difference relatively larger or more reliable for either type of item, and (2) is the personality-achievement difference relatively larger or more reliable for either of the two vocalization groups? Also, are the variabilities any different?

The first issue is dealt with in Tables XVII (a) and XVII (b), the second issue in Tables XVIII (a) and XVIII (b) immediately following.

Table XVII (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO PERSONALITY ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP
(25 in Each Group)

	Average	S.D. (dis)	V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	13.516	12.51	92.55	2.50	7.249	2.79	—
(2)	6.267	5.93	94.62	1.18			2.59 (99)

Table XVII (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN TO ACHIEVEMENT ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP
(25 in Each Group)

	Aver- age	S.D. (dis)	V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	11.808	9.53	80.70	1.91	5.936	2.18	—
(2)	5.872	5.29	90.08	1.06			2.72 (99)

The difference in Table XVII (a) is .73 times the average deflection for all personality items, and the difference in Table XVII (b) is .67 times the average deflection for all achievement items. The reliabilities are both high and are nearly equal. Non-vocal variabilities are higher for both types of items.

It appears, therefore, that the vocalization difference is neither markedly larger nor more reliable for either type of item.

Table XVIII (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS
(25 Subjects)

	Average	S.D. (dis)	V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	13.516	12.51	92.55	2.50	1.708	1.84	.68
(2)	11.808	9.53	80.70	1.91			.92 (82)

Table XVIII (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE DEFLECTIONS GIVEN BY THE NON-VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS
(25 Subjects)

	Aver- age	S.D. (dis)	V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	6.267	5.93	94.62	1.18	.395	.71	.80
(2)	5.872	5.29	90.08	1.06			.55 (71)

The difference in Table XVIII (a) is 13% of the average deflection for all vocal responses, while the difference in Table XVIII (b) is 6.5% of the average deflection for all non-vocal responses. The personality-achievement difference therefore is relatively twice as large for the vocal group as for the non-vocal group; it is also somewhat more reliable for the vocal group than for the non-vocal group, but the difference between the two C.R.'s is small, and both are small in comparison with most of the C.R.'s for the personality-achievement difference. The personality deflections are somewhat more variable in both vocalization groups, the excess over the variability of the achievement deflections being greater in the vocal group. Consequently, it may be concluded that the personality-achievement difference is slightly more pronounced in the vocal group than in the non-vocal group. The direction of the difference,

however, is the same in both groups.

Relationships Between the Sex Difference and the Vocalization Difference. While sex and vocalization were considered separately in relation to their effects on the personality-achievement difference, no data were presented regarding their relationships to each other. There are two reasons for this omission. In the first place, such a comparison has no direct bearing on the main problem. In the second place, the numbers of subjects in the groups which would have to form the basis of such a comparison are too small for reliability computations, since N would be either 10 or 15. Table XII shows that the tendencies for the men and the vocal group to give larger deflections are consistent for these smaller groups, except in comparisons involving the men who did not vocalize their answers. However, the differences involved in the inconsistencies are the smallest differences in the Table, and since they likewise involve the smallest numbers of subjects, they cannot be considered as serious exceptions to the general tendencies.

Mean Deflections for the Favorable, Unfavorable, "?" and Mixed Classes of Items. Through an oversight, the arrangement of the four selected classes of items in the series did not control the positional factor adequately, hence little can be said regarding the mean deflections for the four classes. Table XIX gives the average positions and mean deflections for each of the four classes of items. The Favorable items were accompanied by the largest deflections in both types. In

Table XIX
AVERAGE SERIAL POSITIONS AND MEAN GALVANIC DEFLECTIONS FOR THE FOUR SELECTED CLASSES OF ITEMS

Class	Personality		Achievement	
	Aver. Pos.	Mean Deflec.	Aver. Pos.	Mean Deflec.
Favorable	16.2	12.81	20.0	9.03
Unfavorable	23.3	8.19	20.8	8.11
"?"	22.2	9.60	19.9	8.45
Mixed	20.3	9.16	21.3	8.84
Average	20.5	9.94	20.5	8.61

the personality type, however, the Favorable items averaged nearly four serial positions closer to the beginning than the average for all items. The Unfavorable items also have the lowest mean deflection in both types, but in the personality type, they appear somewhat later in the series.

While no definite conclusions may be drawn from Table XIX, the general nature of the data therein leads to the suspicion that if the positional factor had been more carefully controlled, there would be no considerable differences between the mean deflections for the four selected classes of items within each type, but the difference between the types would hold true for each class. The latter inference is true as the table stands; the former gains some support from the fact that for the achievement items, where the average positions of the four classes vary over a comparatively narrow range, the greatest difference in mean deflections is .92, which is no larger than the average personality-achievement difference which were considered earlier in this chapter. A further inference, therefore, is that the intra-type differences are greater than the inter-type differences because of the sex and vocalization factors, and not because of different qualities within the same type.

General Summary of Chapter IV. While the relatively small number of subjects available for this study has been an admitted handicap in the attempt to demonstrate reliable differences, it is felt that the data concerning the psychogalvanic responses measured in experiment 2 justify the following conclusions, tentatively at least:

- (1) Responses to personality items produce relatively large deflections with greater frequency than the responses to achievement items.
- (2) Responses to personality items produce relatively fewer zero and very small deflections than the responses to achievement items.
- (3) The mean deflection accompanying the responses to personality items is quite consistently larger than the mean deflection accompanying the responses to achievement items, regardless of the classifications on which the comparison is based, provided that the results from all subjects (or all of either sex or vocalization group) are included in the comparison.
- (4) The personality-achievement difference reaches its maximum reliability in the second half of the stimulus series, after the positional effect has been largely exhausted.

(5) A sudden shift from a personality item to an achievement item tends to result in a decreased mean deflection accompanying the responses to the achievement item; while a sudden shift from an achievement item to a personality item tends to result in an increased mean deflection accompanying the responses to the personality item.

(6) The responses made by the men are accompanied by markedly larger mean deflections than the responses made by women. The relative magnitude of the sex difference is approximately the same for both types of items; but the deflections given by the men are more variable for the personality item responses, while the deflections for the women are more variable for the achievement item responses. The personality-achievement difference is perhaps very slightly larger and more reliable for the women than for the men.

(7) Answers made vocally are accompanied by deflections larger than those accompanying answers made non-vocally. The vocalization difference is neither relatively larger nor more reliable for either type of item; but the personality-achievement difference is relatively twice as large and somewhat more reliable for the vocal group.

(8) Excepting where the sexes are being compared, deflections accompanying responses to personality items are more variable than deflections accompanying the responses to achievement items.

(9) The personality-achievement difference is much less consistent for individual subject records than for the group records. Two general statements may be made concerning the correspondence between individual and group results: (a) where the personality-achievement difference is most reliable, small majorities of the subjects follow the group tendency; and (b) in all cases, the excess of mean personality deflection over mean achievement deflection is larger on the average than the excess of mean achievement deflection over mean personality deflection.

- - - - -

Chapter V

REACTION TIMES IN EXPERIMENT 2

The results presented in this chapter are based on the performances of the same 50 subjects whose records were used in Chapter IV. Reliability computations, as before, will be based on individual subject averages.

Distribution of All Reaction Times.

The 50 subjects gave 1992 usable reaction time responses. The mean reaction time (true mean) for all responses was 7.669 seconds. For the two types, the separate mean reaction times were: personality, 7.495 seconds; achievement, 7.842 seconds. The difference agrees with the corresponding difference found in experiment 1, although the experiment 2 reaction times are longer than those measured in experiment 1.

Figure 21 shows the distribution of the reaction time responses for the two types of items. It is rather strikingly similar to

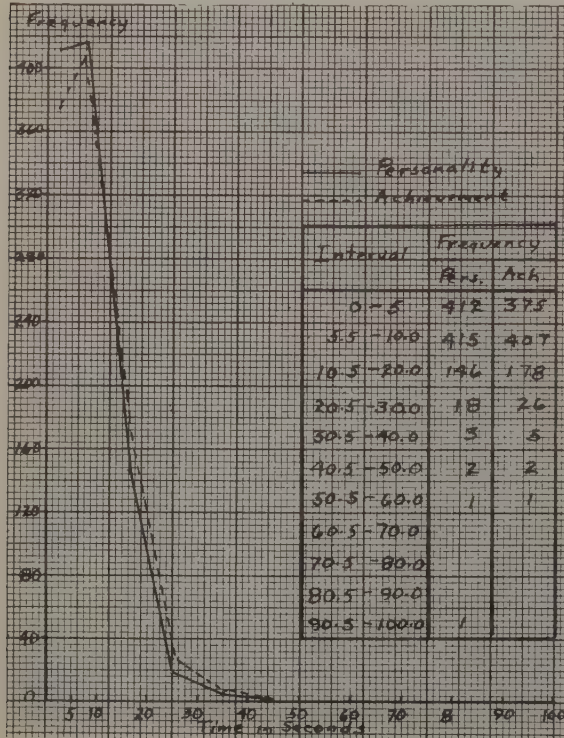


Figure 21. Distribution of All Reaction Times Accompanying the Responses to Personality and Achievement Items. Experiment 2

the corresponding distribution of psychogalvanic responses; that is, there is a high degree of skewness. The two lower intervals (0-10 seconds) have personality frequencies which are higher than the achievement frequencies. The higher intervals have slightly greater achievement frequencies, indicating a tendency for small values to be found less frequently among the achievement reaction times, and vice versa.

Positional Curve of Reaction Times.

Serial position was found to be a major factor influencing the gross magnitude of the psychogalvanic responses in experiment 2. It was therefore deemed desirable to study the effect of serial position on the other measures as well. Figure 22, on the following page, is a positional curve of the reaction times measured in experiment 2. The dotted line is the same curve once smoothed. The general direction of the curve is definitely horizontal. At least, there is no positional tendency which is at all comparable with the sharp decrease in mean deflections seen in Figure 13. The graphic data therefore indicate that there is no consistent or pronounced positional effect influencing the reaction time responses.

Correlational data point to the same conclusion. The correlation between serial position and mean reaction time is $-.17 \pm .10$, which is too small to have any significance. The correlation between the reaction times for corresponding positions on belts 2n and 3n is $.00 \pm .11$, which is further evidence that the mean reaction time has little to do with serial position. On the other hand, the correlation between the mean reaction times for the corresponding items on belts 2n and 3n is $.49 \pm .08$. While this is not a large correlation in itself, it is sufficiently large in comparison with the positional correlations to indicate that the item itself is the chief determiner of the reaction time, rather than the serial position.

Length of Item and Reaction Time. Another factor which might conceivably influence reaction time is the length of the item, the supposition being that a longer question would naturally require more time for reading and grasping the content. "Length of item" is defined arbitrarily as the total number of

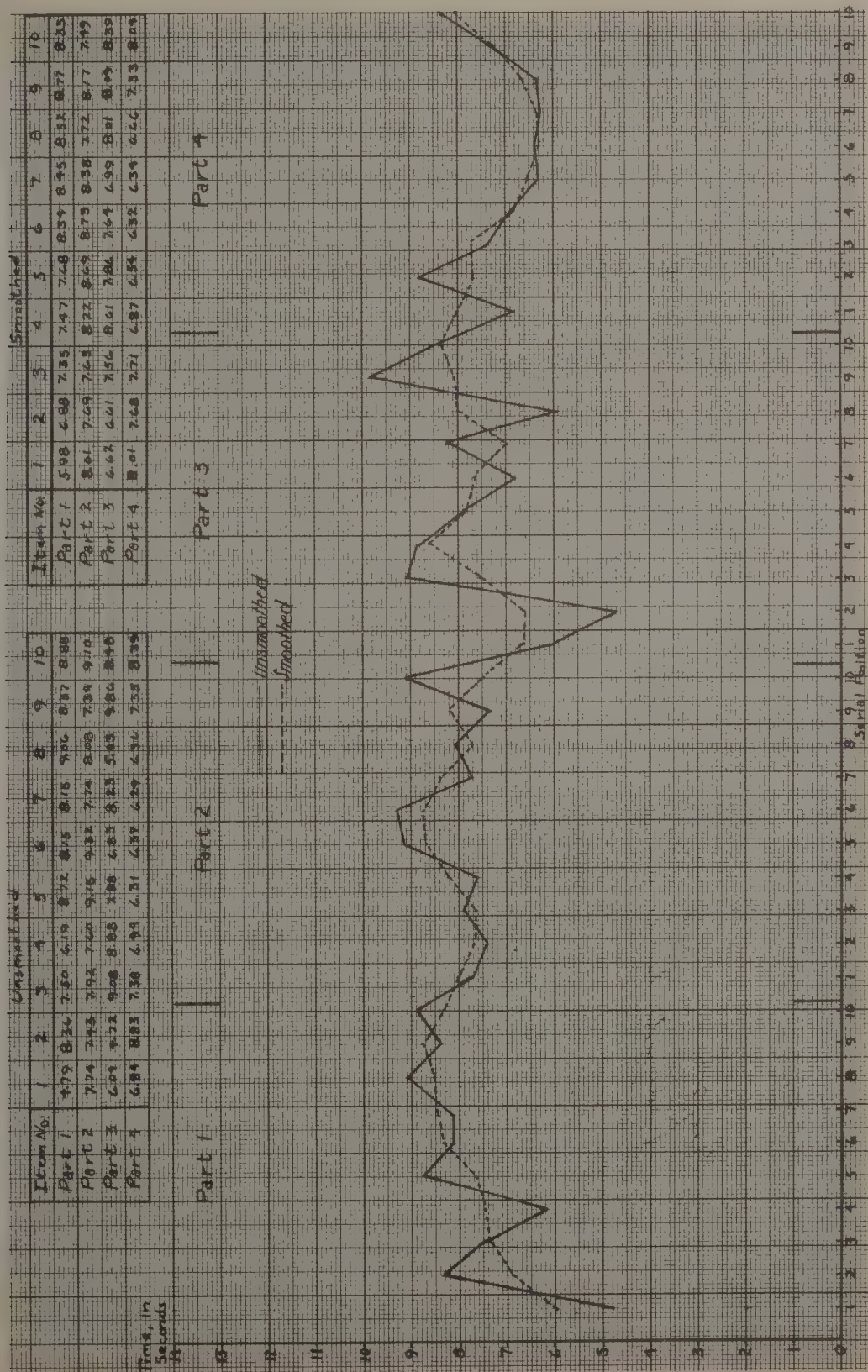


Figure 22. Average Reaction Time for Each Serial Position. Experiment 2

letters in the words making up the item, although it is recognized that this is by no means the sole criterion which might be used, since both the difficulty and familiarity of the words, as well as the complexity of the thought, might well be included under the factors influencing the reading time.

The correlation between length of item as defined above and mean reaction time for each item is .18±.10. Like the positional correlations, this correlation is too small to be of any particular significance. Moreover, the average length for the personality items was 44.4 letters, and for the achievement items, 40.8 letters. It has already been shown that the mean reaction time was shorter for the personality items, the opposite of what might be expected if length were the important factor.

It may therefore be concluded that length of item does not influence the reaction times measured in this study to any appreciable extent.

Consistency of the Personality-Achievement Difference. Table XX gives the mean personality, achievement and total reaction times for several important classifications of items. Sex differences and vocalization differences will be considered in a later Table. Since position is not an important factor, the various Parts of the series will not be considered separately, as was done for the psychogalvanic responses. The values in Table XX are the true means (sum/N) of the ungrouped reaction times, and not the individual subject averages.

Table XX
MEAN PERSONALITY, ACHIEVEMENT AND TOTAL REACTION TIMES
FOR VARIOUS CLASSIFICATIONS OF ITEMS IN EXPERIMENT 2

Classification	Person- ality Mean R.T.	Achieve- ment Mean R.T.	Total Mean R.T.
All items	7.495	7.842	7.669
Belt 2n items	8.072	8.375	8.223
Belt 3n items	6.919	7.316	7.118
Items answered vocally	7.291	7.670	7.474
Items answered non-vocally	7.700	8.014	7.899
Favorable items	7.048	5.994	6.525
Unfavorable items	6.721	8.363	7.542
Mixed items	8.450	9.056	8.731
"?" items	7.760	7.939	7.849

With a single exception (Favorable items), Table XX shows consistently larger mean reaction times for achievement items than for personality items. The personality-achievement difference holds good for all

items, for the two belts, for both vocalization groups, for three of the four selected classes of items and (Table XXI) for the two sexes also. As in the case of the psychogalvanic responses, it is difficult for this sort of consistency to be explained on the basis of chance, and it is likewise difficult to find any constant error to which the difference might be due.

Individual Subject Average Reaction Times. Table XXI gives the individual subject averages classified according to the sex and vocalization groups.

Table XXI
INDIVIDUAL SUBJECT AVERAGE REACTION TIMES FOR THE MEN
AND WOMEN AND FOR THE VOCAL AND NON-VOCAL GROUPS

Classification	Average R.T., Men	Average R.T., Women	Average R.T., All
All responses	8.629	7.026	7.667
All personality responses	8.456	6.872	7.506
All achievement responses	8.823	7.213	7.857
All vocal responses	8.845	6.575	7.484
All non-vocal responses	8.436	7.261	7.878
Vocal personality responses	8.755	6.308	7.286
Vocal achievement responses	8.932	6.848	7.682
Non-vocal personality responses	8.157	7.436	7.724
Non-vocal achievement responses	8.715	7.578	8.033

The personality-achievement difference (larger achievement reaction time) is consistent throughout Table XXI. Likewise, the men give consistently larger reaction times than the women. The evidence regarding a possible vocalization difference is consistent for each sex, but is not the same for both sexes. For the men, the vocal reaction times are longer, while the reverse is true for the women.

Figure 23 shows the distribution of the individual subject reaction times for the personality and achievement types. The frequency of personality reaction times is slightly higher in the two lower intervals, while the frequency of achievement reaction times is slightly higher in the remaining intervals. Figure 23 thus agrees substantially with Figure 22, and in addition is somewhat less skewed.

The Personality-Achievement Difference, All Subjects. Table XXII considers the reliability of the personality achievement difference indicated in Figure 23.

The difference here is somewhat more reliable than the corresponding difference in deflections (Table VII), but again it falls short of complete reliability. The

Table XXII

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT REACTION TIMES TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS (50 Subjects)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff.(diff)	r	C.R.
(1)	7.506	2.44	32.50	.34	.351	.179	.86
(2)	7.857	2.45	31.05	.34			1.95 (97)

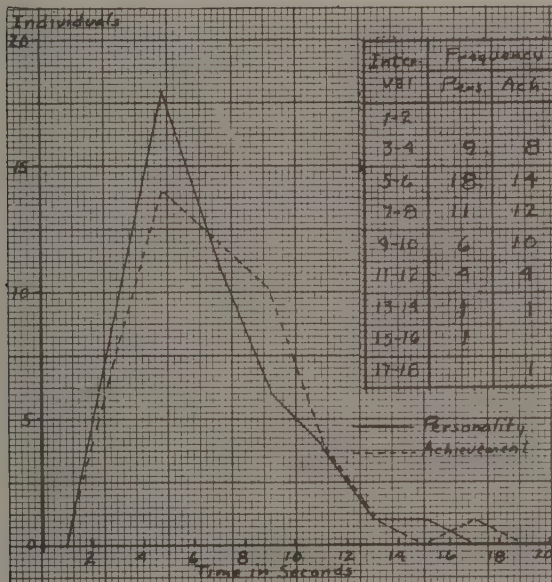


Figure 23. Distribution of Individual Subject Average Reaction Times. Experiment 2

variabilities are practically equal for the two types, and are much smaller than the deflection variabilities, thus showing that the individual differences are likewise much smaller for the reaction times than for the psychogalvanic responses.

The average personality reaction time is exceeded by 26 of the achievement values, while only 18 of the personality values exceed the average achievement reaction time. Larger mean achievement reaction times are given by 31 subjects, larger mean personality reaction times by 19 subjects. Both of these findings somewhat increase the probability that the personality-achievement difference is real.

The Sex Difference, All Reaction Times. Table XXIII presents the data concerning the apparently longer reaction times given by the men as compared with those for the women.

Table XXIII

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT REACTION TIMES GIVEN BY (1) THE MEN, (2) AND THE WOMEN (20 Men, 30 Women)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff.(diff)	r	C.R.
(1)	8.629	2.44	28.27	.54	1.603	.70	—
(2)	7.026	2.47	35.15	.45			2.29 (99)

Figure 24 shows the same data graphically. There is a fairly sharp separation between the distributions for the men and for the women. The two lower intervals contain over four times as many women as men (percents) and the three higher intervals contain nearly twice as many men as women. Only 23.3% of the women exceed the average of the men, while 75% of the men exceed the average of the women. These facts in combination with the data from Table XXIII seem to justify the conclusion that the longer reaction time for the men represents a real difference in spite of the lack of complete statistical reliability.

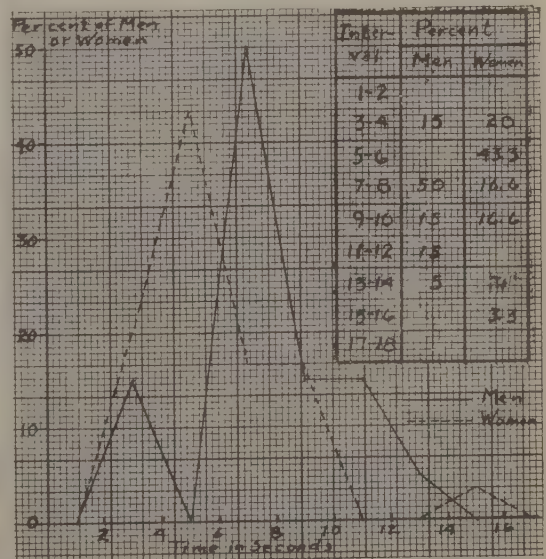


Figure 24. Distribution of Individual Average Reaction Times for Men and Women. Experiment 2

The Vocalization Difference, All Reaction Times. The distributions of the reaction times for the vocal and non-vocal groups are given in Figure 25 and the reliability of the difference is treated in Table XXIV.

Table XXIV

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES FOR (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP (25 in each Group)

	Aver- age	S.D. (dis)	V	S.D. (aver)	Diff. (diff)	r	C.R.
(1)	7.484	3.13	41.82	.602	.394	.70	.56
(2)	7.878	1.85	23.48	.37			(71)

Figure 25 in itself does not lead to the expectation of a very reliable difference between the vocal and non-vocal reaction times, since the vocal reaction times are in excess at both ends of the distribution, rather than being more numerous at one end and less numerous at the other end. Also, it was seen from Table XXI that the direction of the vocalization difference is not the same for the two sexes.

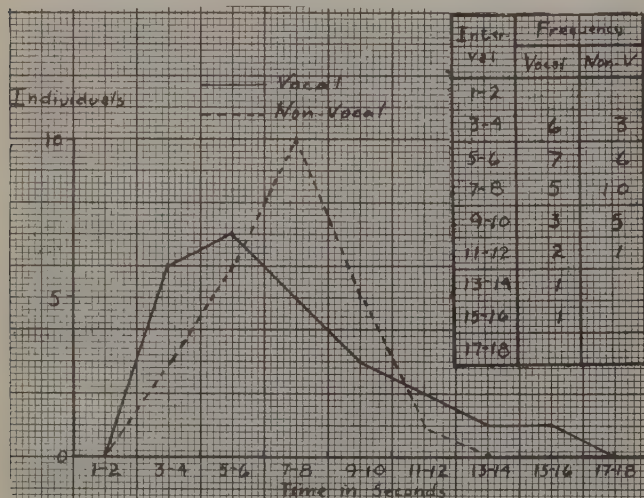


Figure 25. Individual Subject Reaction Times for the Vocal and Non-vocal Groups. Experiment 2

Table XXIV shows the expected low reliability. It may therefore be concluded that for the group as a whole, vocalization of answers is not a factor operating consistently on the magnitude of the reaction times. The vocal reaction times seem to be considerably more variable, however, than the non-vocal reaction times. Comparison of the opposite differences for the two sexes would not be germane to the problem.

Relationships Between the Sex Difference and the Personality-Achievement Difference. The issues to be raised here are (1) Whether the personality-achievement difference is relatively larger or more reliable

for either sex, and (2) whether the sex difference is relatively larger or more reliable for either type of item. Variabilities will also be considered. Tables XXV (a) and XXV (b) give the data concerning the first issue.

Table XXV (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE MEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS (20 Subjects)

	Aver- age	S.D. (dis)	V	S.D. (aver)	Diff. (diff)	r	C.R.
(1)	8.456	2.40	28.38	.53	.367	.84	1.15
(2)	8.823	2.61	29.58	.58			(87)

The difference of .367 for the men is 4.25% of the average reaction time for men on all responses, while the corresponding difference of .341 for the women is 4.85% of their total average reaction time. The personality-achievement difference is therefore 1.14 times as large for the women as for the men. The same difference is apparently more reliable for the women as shown by the C.R.'s; but the smaller number of men renders this conclusion uncertain. Reaction times given by the women are more variable than those given by the men for both types of items, but the differences between the personality V's and the achievement V's for both men and women are too small to have any particular significance.

The second issue is considered in Tables XXVI (a) and XXVI (b). The sex difference in Table XXVI (a) is 21% of the average personality reaction time, while the sex difference in Table XXVI (b) is 20% of the average achievement reaction time. The C.R.'s

Table XXV (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE WOMEN TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS (30 Subjects)

	Aver- age	S.D. (dis)	V	S.D. (aver)	Diff. (diff)	r	C.R.
(1)	6.872	2.40	34.92	.44	.341	.95	2.25
(2)	7.213	2.66	36.97	.48			(99)

Table XXVI (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE PERSONALITY ITEMS BY (1) THE MEN, AND (2) THE WOMEN
(20 Men, 30 Women)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	8.456	2.40	28.38	.53	1.584	.69	2.29
(2)	6.872	2.40	34.92	.44			(99)

Table XXVI (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE ACHIEVEMENT ITEMS BY (1) THE MEN, AND (2) THE WOMEN
(20 Men, 30 Women)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	8.823	2.61	29.58	.58	1.610	.75	2.14
(2)	7.213	2.66	36.97	.48			(98)

are likewise nearly equal. It must be concluded therefore that the sex difference is neither larger nor more reliable for either type of item.

Relationships Between the Vocalization Difference and the Personality-Achievement Difference. The difference between the average reaction times of the vocal and non-vocal groups for all items was found to be very unreliable (Table XXIV). Next come the questions: whether (1) the personality-achievement difference is larger or more reliable for either of the vocalization groups, and whether (2) the vocalization difference is larger or more reliable for either type of item. Tables XXVII (a) and XXVII (b) present the data concerning the first issue.

Table XXVII (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS
(25 Subjects)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	7.286	2.98	40.90	.596	.396	.44	.76
(2)	7.682	3.44	44.76	.686			(82)

Table XXVII (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN BY THE NON-VOCAL GROUP TO (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS
(25 Subjects)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	7.724	2.03	26.28	.406	.309	.374	.55
(2)	8.033	1.99	24.77	.398			(79)

The difference for the vocal group is 5.29% of the total average vocal reaction time, while the difference for the non-vocal group is 3.92% of the total average non-vocal reaction time. The personality-achievement difference is therefore 1.33 times as large in the vocal group as in the non-vocal group. The reliabilities are practically equal. Vocal reaction times are considerably more variable than the non-vocal reaction times for both classes of items.

Tables XXVIII (a) and XXVIII (b) take up the second issue.

Table XXVIII (a)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE PERSONALITY ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP
(25 in each Group)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	7.286	2.98	40.90	.596	.438	.71	.61
(2)	7.724	2.03	26.28	.406			(73)

Table XXVIII (b)

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL SUBJECT AVERAGE REACTION TIMES GIVEN TO THE ACHIEVEMENT ITEMS BY (1) THE VOCAL GROUP, AND (2) THE NON-VOCAL GROUP
(25 in each Group)

	Aver- age	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	7.682	3.44	44.76	.686	.351	.78	.45
(2)	8.033	1.99	24.77	.398			(67)

The difference in Table XXVIII (a) is 5.8% of the total average personality reaction time, while the difference in Table XXVIII (b) is 4.4% of the total average achievement reaction time. The personality

type therefore has a slightly larger and slightly more reliable vocalization difference than the achievement type.

Reaction Times vs. Psychogalvanic Responses. The data in this section are based on correlations.

For the individual subject averages, the correlation between reaction time and psychogalvanic response, all items, was $-.27 \pm .087$. For the personality items only, the correlation was $.14 \pm .093$, and for the achievement items only it was $.23 \pm .090$. These correlations are small, they are inconsistent, and they have large probable errors.

For the individual item averages, the correlation between psychogalvanic response and reaction time was $-.09 \pm .106$, while the scatters indicated that the correlations for the separate types would be no more significant.

It may be concluded, therefore, that the data show no significant relationships between reaction times and psychogalvanic responses, either for the individual subjects or for the individual items.

Reaction Times for the Four Selected Classes of Items. The data involved in this discussion have already been given in Table XX, but for convenience they are repeated in Table XXIX.

Table XXIX
MEAN PERSONALITY, ACHIEVEMENT AND TOTAL REACTION
TIMES FOR THE FOUR CLASSES OF ITEMS SELECTED
ON THE BASIS OF THE GROUP TEST RESULTS

Class	Personality Mean R.T.	Achievement Mean R.T.	Total Mean R.T.
Favorable	7.048	5.994	6.525
Unfavorable	6.721	8.363	7.542
Mixed	8.450	9.056	8.731
"?"	7.760	7.939	7.849

Inasmuch as there is no positional factor to interfere with the interpretation of the reaction times, these results may be discussed with somewhat greater assurance than was the case with the corresponding deflection results.

As already noted, the Favorable items are the only classification in which the mean achievement reaction time failed to exceed the mean personality reaction time. This fact, however, does not constitute a serious exception to the general tendency for achievement reaction times to be longer. The favorable items included the easiest questions in the entire achievement list, and it is not at all surprising to find that they were answered in

less time than were the items in any other class. On the other hand, the longer favorable personality reaction time might easily be interpreted as the result of deliberately seeking for the "right" answer. The sharply increased mean reaction time for the unfavorable achievement items reflects their greater difficulty. If the favorable and unfavorable classes of each type are taken as a single unit, the average is still higher for the achievement type.

The mixed class has the longest reaction time for both types, indicating that for the group as a whole, the lack of agreement on the answer is paralleled by greater hesitation in answering. It was originally anticipated that the "?" class would have the longest average reaction time. Since it does not, one is led to suspect that the "?" answer, so far as the achievement items are concerned, may represent the lack of any real effort at all. (It will be shown in the next chapter that this class also has the lowest memory value of all.) This is probably less true in the personality type, where the "?" class has the next to the longest mean reaction time.

On the whole, the data in Table XXIX would seem to justify the tentative conclusions that: (a) the choice between favorable and unfavorable answers affects the reaction times for the two types differently; (b) the mixed items provoke the greatest hesitation in both types; and (c) the mean reaction times for the selected classes lend general confirmation to the observed tendency for achievement reaction times to exceed personality reaction times.

General Summary of Chapter V. Neither serial position nor length of item is significantly related to reaction time. There are consistent tendencies for achievement reaction times to be longer than personality reaction times, and for the men to have longer reaction times than the women. Vocalization seems to influence reaction time, but it is complicated by sex factors which were not investigated. Variabilities are nearly equal for the two types, but are greater for the women and the vocal group. The personality-achievement difference is somewhat larger for the women and the vocal group, but neither sex nor vocalization affects the reliability of the difference markedly. Reaction times are not correlated with the magnitudes of the psychogalvanic responses accompanying them. The average reaction times for the four selected classes show different type tendencies in the Favorable and Unfavorable classes, but confirm the general personality-achievement difference.

Chapter VI

IMMEDIATE RECALL, JUDGMENTS OF "DIFFICULT", AND CHANGED ANSWERS, IN EXPERIMENT 2

While dealing primarily with the memory-score results from experiment 2, this chapter will consider briefly the data concerning the judgments of "difficult" made by the subjects, and will also include a brief note regarding the few changes of answer which were made.

General Results. The 50 subjects of experiment 2 recalled a total of 675 items, an average of 13.500 items per subject. The number of personality items recalled was 358, an average of 7.160 per subject. The number of achievement items recalled was 317, or 6.340 per subject. The women recalled an average of 14.233 items, the men an average of 12.400 items. The subjects in the vocal group recalled an average of 13.440 items, those in the non-vocal group 13.560 items. Table XXX gives the individual subject average memory scores classified to show the personality achievement, sex and vocalization differences.

Effect of Serial Position on Memory for Items. Serial position is a potential factor influencing the memory scores since for each subject there would be tendencies due to primacy and recency. It might be expected that the items first presented in the series and those coming near the end of the series would be remembered somewhat more easily than those items occurring in the middle of the series. The results tend to confirm the expectation, since the average frequency of recall per item in each of the four Parts of the series was: Part 1, 16.8; Part 2, 14.6; Part 3, 16.2; and Part 4, 20.9. The items in Part 4 were thus favored over all other items, Part 1 has a small advantage over Part 3, and Part 2 has the lowest average recall per item of any of the Parts. Figure 26 is a positional curve of the recall scores. The correlation between position and recall score is $.24 \pm .10$, also indicating a very slight tendency for the subjects to recall the items occurring late in the series more often than those occurring early in the series.

While the positional effect is therefore present, it can be demonstrated that it is not a decisive factor influencing the com-

parisons in which the study is interested. The alternation of the order of items within the belts and for the two belts causes the positional effect to cancel out when the results for both belts are combined. The correlation between the average position and the average memory score for the 40 items is $-.04 \pm .106$, and the correlation between the memory scores for the corresponding positions on the two belts is $.25 \pm .10$. On the other hand, the correlation between the memory scores for the 40 items on the two belts is $.58 \pm .07$. That is, in spite of the differences in serial position, the memory scores for the 40 items agree much closer with each other than would be expected if serial position were the chief determining factor. The item itself, and not its position in the series, thus seems to be the chief factor influencing recall.

Type, Sex, and Vocalization Difference in Recall. The individual subject average memory scores for various classifications are given in Table XXX.

Table XXX
INDIVIDUAL SUBJECT AVERAGE MEMORY SCORES CLASSIFIED TO SHOW DIFFERENCES DUE TO TYPE OF ITEM, SEX, AND VOCALIZATION OF ANSWERS

Classification	Average number of times recalled by:		
	Men	Women	Both
All items	12.400	14.233	13.500
All personality items	6.200	7.800	7.160
All achievement items	6.200	6.433	6.340
All items answered vocally	12.100	14.333	13.440
All items answered non-vocally	12.700	14.133	13.560
Vocal personality items	5.800	8.200	7.240
Vocal achievement items	6.300	6.133	6.200
Non-vocal personality items	6.600	7.400	7.080
Non-vocal achievement items	6.100	6.733	6.480

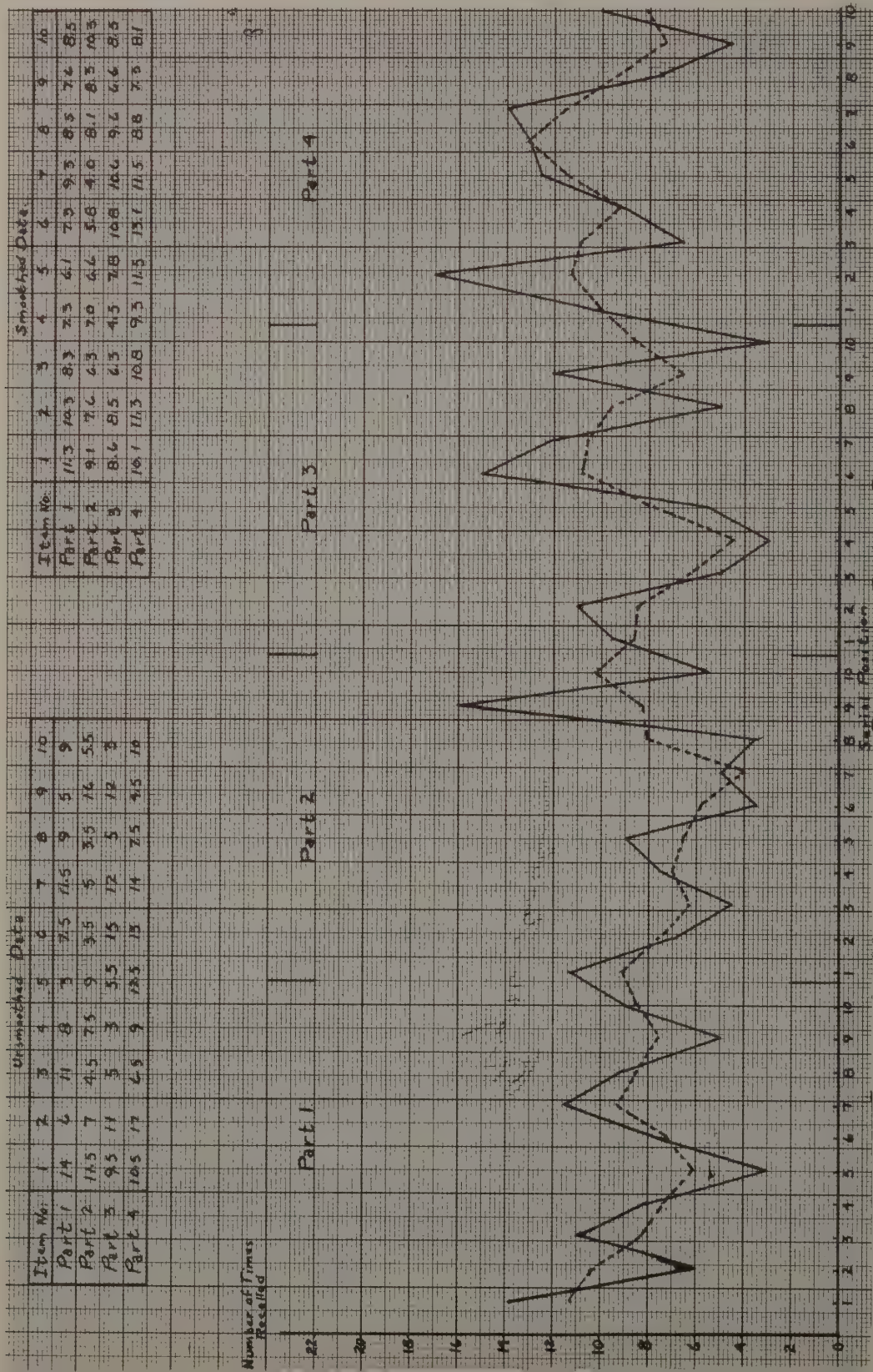


Figure 26. Average Memory Value of Each Serial Position.
Experiment 2

Table XXX indicates that the women regularly recall more personality items than achievement items. This is also true of the men who did not vocalize their answers. The men who did vocalize their answers recalled more achievement items, thus making the two types equal in recall value for the men as a group. The women regularly recall more items than the men, except in the vocal achievement classification. The vocalization difference for all subjects is negligible, and it is opposite in direction for the sexes taken separately, being also relatively larger for the men. The personality-achievement difference is larger for the vocal group than for the non-vocal group.

Average Recall Scores for the Selected Classes of Items. Table XXXI gives the data concerning the recall of the four selected classes of items in each type.

Table XXXI
AVERAGE FREQUENCY OF RECALL OF ITEMS IN EACH
OF THE FOUR SELECTED CLASSES IN EACH TYPE

Class	Frequency of Recall		
	Personality	Achievement	Both
Favorable	6.80	8.50	7.65
Unfavorable	14.20	8.20	11.20
"?"	7.00	5.20	6.10
Mixed	8.40	10.20	9.30
Average	9.10	8.02	8.56

Table XXXI shows that the superior memory value of the personality items is not distributed evenly, but is concentrated largely in the Unfavorable class. In general, the four classes behave quite differently for the two types. In the personality type, the Favorable and Unfavorable items have the lowest and highest memory values respectively; while in the achievement type both of these classes occupy intermediate positions in the scale of memory values, and the high and low extremes are found in the Mixed and "?" classes. It is therefore probable that, whatever be their exact nature (affective or otherwise), the subjective conditions which determine the memory values of test items such as these operate somewhat differently on the Favorable and Unfavorable classes in the two types. It is not difficult to find at least one plausible explanation of this difference. From the subject's point of view, the Unfavorable answer to an achievement item is rarely accompanied by the realization that it is incorrect; to him, the answer is usually correct, or at least it represents his best judgment concerning the correct answer. On the other hand, the intelligent adult can hardly escape

being aware of the fact that one of the answers to a personality item indicates a condition which is less desirable, socially or morally, than the condition indicated by the other answer. This being the case, it is easy to understand how the consciously "wrong" answer, if given to a personality item, would tend to be impressed more vigorously in the memory than some other answers. In the achievement type of item, the subject usually makes no such conscious distinction; therefore there is no reason why the Unfavorable achievement item should be noted any more vividly than the Favorable item.

The Mixed and "?" class memory values are also of interest. It was observed in the preceding chapter that in both types the Mixed class had the highest average reaction times of any of the classes. Here the Mixed classes are highest and second highest in memory value for the achievement and personality types respectively. This coincidence of high reaction times and high memory values for the Mixed items is somewhat unexpected in view of the fact that the correlation between mean reaction time and total memory score for all 40 items is $-.21$. Evidently the indecisiveness of the group as a whole with respect to these Mixed items also acts to increase their memory values. The "?" items have the lowest memory values for the whole group of items and in the achievement type as well, while they are next to the lowest in the personality type.

The Personality-Achievement Difference in Memory Scores for All Subjects. Table XXXII considers the evidence regarding the apparently superior memory value of the personality items for all subjects.

Table XXXII
DATA CONCERNED WITH THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE INDIVIDUAL SUBJECT MEMORY SCORES FOR (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS

	Average	S.D. (dis)	S.D. (aver)	S.D. Diff.	(diff)	r	C.R.
(1)	7.160	2.28	71.84	.32	.820	.31	2.64
(2)	6.340	2.35	37.06	.33		.53	(99)

The C.R. (2.64) is nearly the required three times its S.D., and is more reliable than either of the corresponding differences in personality and achievement deflections and reaction times. The achievement memory scores are 1.16 times as variable as the personality memory scores. Figure 27 makes the same comparison graphically. The three lowest intervals contain nearly twice as many achievement memory scores as personality memory scores,

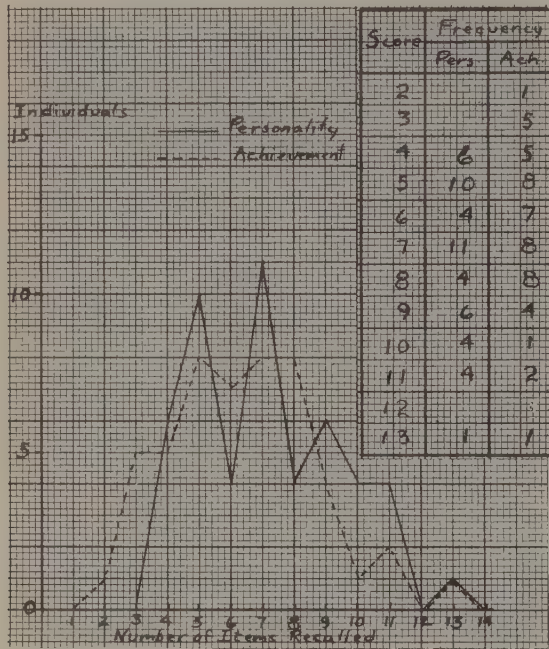


Figure 27. Distribution of Individual Subject Memory Scores

Experiment 2

as personality memory scores, while the reverse is true of the four highest intervals. Of the individual subjects, 27 agree with the group tendency (rather poorer agreement than was found in experiment 1, where 14 out of 20 subjects recalled more personality items, but the preponderance of women in experiment 1 may explain this in part), 18 disagree, and for 5 the scores were equal.¹ For those who agreed with the group tendency, the average excess of personality memory score over achievement memory score was 2.66; while for those who disagreed with the group tendency, the average excess of achievement memory score over personality memory score was 1.73. In terms of overlapping, 30 of the personality memory scores exceed the mean achievement memory score, while only 16 of the achievement memory scores exceed the mean personality memory score. Taking the group as a whole, therefore, the personality-achievement difference is probably a real one, although there are important sex influences, as will be demonstrated presently.

The Sex Difference in Total Memory Scores. Figure 28, shows the distribution of the total memory scores for the men and the women. (The ordinates are percentages of the

groups.) The four lower intervals contain 80% of the scores made by the men, compared with 50% of the scores made by the women. In the four higher intervals, the situation is reversed, 20% of the men and 50% of the women being found there. Larger scores are thus more frequent with the women, and vice versa. Only 4 men exceed the mean score for the women, while 17 women exceed the mean score for the men. (The corresponding percentages are 25 and 56.6.) Table XXXIII gives the reliability data for the sex difference.

Table XXXIII

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE INDIVIDUAL AVERAGE TOTAL MEMORY SCORES FOR (1) THE MEN, AND (2) THE WOMEN.
(20 Men, 30 Women)

	Average	S.D. (dis)	V	S.D. (aver)	S.D. Diff (diff)	r	C.R.
(1) 12.400	3.26	26.28	.75		1.833	1.05	1.74
(2) 14.233	4.19	29.43	.76				(.96)

It is probable that the tendency for the women to recall a larger total number of items than the men is a real difference. The scores for the women are 1.12 times as variable as those for the men.

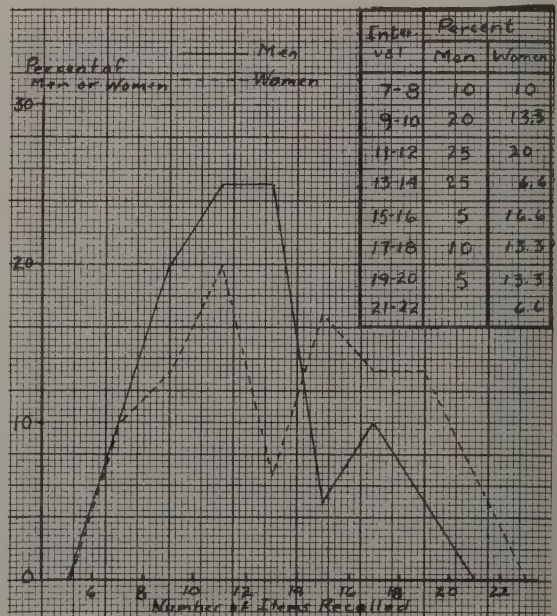


Figure 28. Total Memory Scores of Men and Women
Experiment 2

¹The three subjects of belt 1n also had a single recall quite comparable with the recall of experiment 2; indeed, the results might well have been included with those of experiment 2. All three had higher personality memory scores. Of all subjects tested, 44 had higher personality memory scores, 21 lower, and 8 the same.

Relationships Between the Sex Difference and the Personality-Achievement Difference. Table XXX shows that, while the women score higher than the men in both types of items, the difference in the achievement type is very small (.233) compared with the difference in the personality type. The Table also shows that the average memory scores for the two types of items are equal for the men, but differ by 1.367 for the women. These figures indicate that the sex difference is confined largely to the personality type, and that the personality-achievement difference is due largely to the performance of the women. Tables XXXIV and XXXV present the evidence in support of these conclusions.

Table XXXIV

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE NUMBER OF PERSONALITY ITEMS RECALLED BY (1) THE MEN, AND (2) THE WOMEN.

(20 Men, 30 Women)

	Average	S.D. (dis)	S.D. V	S.D. (aver)	Diff. (diff)	r	C.R.
(1) 6.200	1.72	27.74	.38				
(2) 7.800	2.26	28.97	.41	1.600	.56	—	2.85
							(99)

Table XXXV

DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE INDIVIDUAL MEMORY SCORES MADE BY THE WOMEN FOR (1) PERSONALITY ITEMS, AND (2) ACHIEVEMENT ITEMS.

(30 Subjects)

	Average	S.D. (dis)	S.D. V	S.D. (aver)	Diff. (diff)	r	C.R.
(1) 7.800	2.26	28.97	.41				
(2) 6.433	2.40	37.30	.44	1.367	.41	.53	3.33
							(100)

The C.R. of 2.85 found in Table XXXIV may be considered as strongly indicative of reliability, since the smaller number of men has consistently acted to deflate the sex-difference reliabilities. In terms of overlapping, 70% of the women have higher personality scores than the average score of men, while only 15% of the men have scores higher than the average for the women. Figure 29 shows the same difference graphically. The percentage of men is distinctly higher in the two lower intervals, while in the remaining intervals the women far outnumber the men.

Since the personality-achievement difference seems to be concentrated in the scores for the women, it might be expected to reach its maximum reliability in that group. Figure 30 and Table XXXV present the data. The

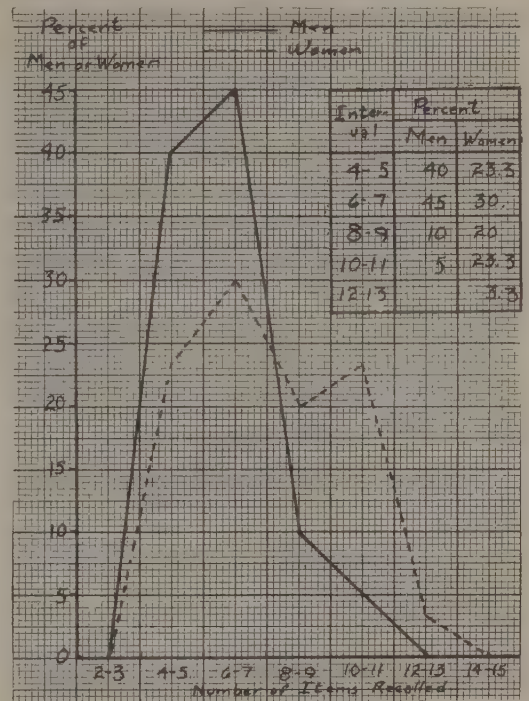


Figure 29. Distribution of Personality Item Memory Scores for Men and Women
Experiment 2

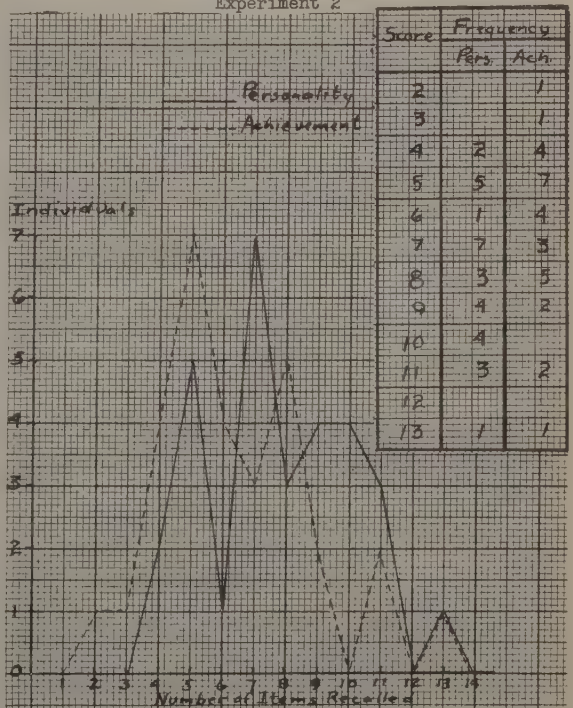


Figure 30. Distribution of Personality and Achievement Item Memory Scores for Women
Experiment 2

difference is completely reliable (i.e., the C.R. is larger than 3.0).

It is concluded, therefore, that the personality-achievement difference in memory scores is closely related to sex, since the difference is due largely to the women's superior memory for personality items.

Relationships Between Vocalization and the Personality-Achievement Difference. It has already been noted in Table XXX that the difference in total number of items recalled by the vocal and non-vocal groups was very small (.120). For the personality items only the difference is .160, and for the achievement items only, .240, and the differences are opposite in direction. Hence the vocalization difference is not significant for either type of item.

On the other hand, the personality-achievement difference for the vocal group (1.04) is nearly twice as large, absolutely, as the corresponding difference (.60) for the non-vocal group. Tables XXXVI(a) and XXXVI(b) consider the data relating to these differences.

Table XXXVI(a)
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE NUMBERS OF (1) PERSONALITY ITEMS, AND (2) ACHIEVEMENTS ITEMS RECALLED BY THE VOCAL GROUP
(25 Subjects)

	Aver- age	S.D. (dis)	V	S.D. (aver)	Diff.	S.D. (diff)	r	C.R.
(1)	7.240	2.27	31.35	.45	1.04	.52	.27	2.00
(2)	6.200	2.13	34.35	.42				(98)

The difference in Table XXXVI(a) is 7.7% of the average memory score for the vocal group as a whole, while the corresponding difference in Table XXXVI(b) is 4.4% of the average memory score for the non-vocal group as a whole. The personality difference is

Table XXXVI(b)
DATA CONCERNING THE RELIABILITY OF THE DIFFERENCE BETWEEN THE AVERAGE NUMBERS OF (1) PERSONALITY ITEMS, AND (2) ACHIEVEMENT ITEMS RECALLED BY THE NON-VOCAL GROUP
(25 Subjects)

	Aver- age	S.D. (dis)	V	S.D. (aver)	Diff.	S.D. (diff)	r	C.R.
(1)	7.080	2.15	30.36	.43	.60	.52	.40	1.15
(2)	6.480	2.54	39.19	.51				(87)

thus considerably larger for the vocal group, and also (as shown by the C.R.'s) somewhat more reliable. In the vocal group, however, there are opposite sex tendencies. The effects of vocalization are thus somewhat inconsistent, on the whole.

Memory Scores vs. Psychogalvanic Responses. There are three principal issues under this topic: first, what is the correlation between mean psychogalvanic response and total memory score for the various subjects; second, what is the correlation for the various items; and third, how does the average deflection for the recalled items compare with the average deflection for all items?

The correlation between total memory score and average deflection for the 50 subjects was $.04 \pm .09$, and the scatter diagrams indicated a corresponding lack of correlation for the separate types of items.

The correlation between average deflection and number of times recalled for the 40 items was $.25 \pm .10$, the scatter diagrams indicating again that the correlations for the separate types would be no better, perhaps worse. Hence we can say that memory scores and mean deflections are not significantly correlated either for the subjects or for the items.

The grand average of the deflections for all items was 9.33, and for the recalled items alone, 9.79. For the personality items alone the corresponding figures were 9.90 and 9.65, and for the achievement items alone, 8.77 and 9.94. The differences are opposite in direction for the two types, and only the achievement difference is large enough to consider seriously. However, the lack of correlation, as indicated above, makes the interpretation of this difference uncertain. In consequence, it must be concluded that there are no marked relationships between memory and pgr.

Memory Scores vs. Reaction Times. It was noted on an earlier page that the correlation between mean reaction time and memory score for the 40 items was $-.21$, with no reason for believing that the correlations for the separate types of items would be any higher. Computed on the basis of the individual subject averages, the correlation between reaction time and memory score is $-.09$. It does not appear, therefore, that there is any marked relationship between these two measures.

Significance of the Lack of Interrelationships. The data presented above and in the corresponding section of Chapter V indicate quite clearly that different groups of subjects are responsible for each of the three

principal tendencies (larger mean personality deflection, larger mean achievement reaction time, and higher personality memory value) which have been suggested by the results of this investigation.

It has been demonstrated that each tendency, considered separately, is followed by at least half of the subjects; and that the mean differences for the contrary cases are, on the average, smaller than the corresponding differences for the agreeing cases. It is also of interest to note the frequency of combinations of agreements with the general tendencies. Only 4 subjects failed to follow at least one of the three tendencies, 19 followed only one, 13 followed two, and 14 subjects agreed with all three of the group tendencies. In terms of percentages, 92% of the subjects agree with at least one of the group tendencies, 54% agree with at least two of the group tendencies, and 28% agree with all three of the group tendencies.

It must therefore be concluded that the three principal measures which have been used are not inter-related, but are affected independently, if at all, by the two sorts of stimuli (personality and achievement items). With respect to each of the measures the influences of individual differences, sex, and the vocalization of answers are considerably more potent in producing gross differences than is the qualitative nature of the stimulus.

Judgments of Difficulty. It will be remembered that the subject was asked to press the answer key twice if he considered the question relatively difficult. A total of 278 such judgments were expressed by the 50 subjects used in the second experiment. Of these 278 judgments, 124 (or 44.6%) accompanied responses to personality items, while 154 (or 55.4%) accompanied responses to achievement items. Each personality item was therefore adjudged difficult 6.20 times, while each achievement item was adjudged difficult 7.70 times, on the average. The individual subject averages are considered in Table XXXVII.

Table XXXVII

RELIABILITY DATA CONCERNING THE DIFFERENCE BETWEEN THE AVERAGE NUMBER OF "DIFFICULT" JUDGMENTS PER SUBJECT FOR (1) THE PERSONALITY ITEMS, AND (2) THE ACHIEVEMENT ITEMS.
(50 Subjects)

	Average	S.D. (dis)	S.D. V	S.D. (aver)	S.D. Diff. (diff)	r	C.R.
(1)	2.48	2.53	102.01	.35			
(2)	3.08	3.15	102.27	.44	.60	.42	.44 1.42
							(92)

Figure 31 shows the individual subject data in graphic form. The lower intervals (fewer judgments of difficulty) have slightly larger personality frequencies, and vice versa.

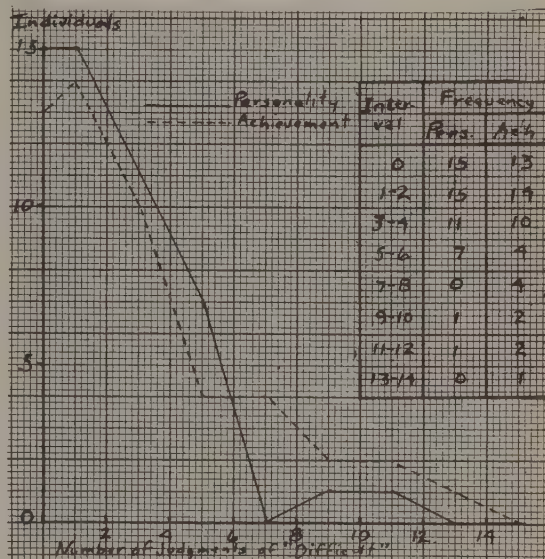


Figure 31. Distribution of Judgments of "Difficult" Accompanying the Responses to Personality and Achievement Items

The evidence thus agrees that the number of judgments of difficult is smaller for the personality items, although the difference is not highly reliable. It harmonizes, however, with the previously found tendencies for the personality item responses to be more rapid, to employ the "?" answer less often, and to be less subject to change.

Tabulation of the "Difficult" judgments for each item and for all subjects in both experiments (Table III-A, in the Appendix) shows that the first four and the sixth achievement items receiving the highest "Difficult" scores were of the "?" class, and that the remaining five out of the ten highest were of the Mixed class. On the other hand, four of the five achievement items with the lowest scores were of the Favorable class. By the way of contrast, the ten personality items with the highest "Difficulty" scores also included four "?" items and three Mixed items, but the two highest were both Favorable items, and the sixth was Unfavorable. Three of the five lowest were Favorable and two were Mixed. It would therefore seem that a judgment of "Difficult" means something different for the two types of items, although the data do not permit any theories as to the nature of the difference.

Changed Answers in Experiment 2. The group-testing results suggested that fewer changes were made in the answers to personality items than were made in the answers to achievement items. The data from experiment 2 point in the same direction, although the number of changes actually made was so small in comparison with the number of subjects that the interpretation thereof is by no means certain.

The 50 subjects of experiment 2 changed only 40 answers or less than 1 per subject. Sixteen of the changes were associated with personality items and 24 with achievement items. Data regarding the accompaniments of the changed answers are given in Table XXXVIII.

Table XXXVIII
MEASURES ACCOMPANYING CHANGED ANSWERS, EXPERIMENT 2

Type	Number of Changes	% Judged Difficult	% Recalled	Mean Reaction Time	Mean Deflection
Pers.	16	18.7	57.5	27.9	25.2
Ach.	24	8.3	57.5	21.6	15.8

It may be seen from Table XXXVIII that the changed answers were accompanied by reaction times and deflections considerably larger than those accompanying all responses, and that they were recalled equally well in both types. Judgments of "Difficult" accompany changed answers somewhat more often in the case of personality items, but the small number of changes makes this distinction dubious. The data are too meager to form the basis for type-differentiation on any very extensive scale.

Summary of Chapter VI. Serial position is a factor in the production of immediate recall of test items, but the alternation of items in the two belts neutralizes

the effect of serial position on the type differences. Women recall more items of both types than men, and the women also recall more personality items than achievement items, the latter difference being completely reliable. Memory scores for the four selected classes of items show decided type differences. The personality-achievement difference is larger and more reliable for the vocal group than for the non-vocal group, but the difference in total memory scores for the two groups is negligible, and the effect of vocalization on the personality-achievement difference is complicated by a sex factor for which the data do not permit adequate investigation. The net effect of vocalization is therefore uncertain. Memory scores are not related significantly either to psychogalvanic responses or to reaction times. The lack of inter-relationships between the three major differentiating agents used in this study indicates that the three main tendencies are produced by different groups of subjects, although 14 of the subjects agree with all three of the group tendencies.

Judgments of "Difficult" tend to be associated more frequently with achievement item responses than with personality item responses, and to show type differences in the classes of items most frequently judged difficult.

Fewer changes of answer are made for the personality items than for the achievement items. Changed answers to both types are characterized by unusually large reaction times and psychogalvanic responses, but the number of changes made was too small to permit consideration of type differences with respect to the latter measures.

- - - - -

Chapter VII

SUMMARY AND CONCLUSIONS

Problem. The purpose of this investigation was to determine whether the responses of a group of subjects to 20 selected personality test items and 20 selected achievement test items would show differences with respect to (a) the psychogalvanic changes accompanying the responses, (b) the reaction times (number of seconds required for giving the response), (c) the memory values (measured by immediate recall of the items to which responses had been made), and (d) certain other measures of lesser importance. The intention of this investigation was to make a contribution to the larger issue concerning the validity of ordinary "objective" testing techniques as applied to the field of personality measurement, through the determination of whether the subjective conditions influencing the responses to personality test items could be differentiated (by means of the above named measures) from the subjective conditions influencing the responses to achievement test items.

Procedure. The 40 test items used in this investigation were selected on the basis of the results obtained by administering the Bernreuter Personality Inventory and a specially constructed General Information Test to 88 upperclass college students. The selected items were divided into four classes denoted respectively Favorable (high frequency of normal or correct answers), Unfavorable (high frequency of neurotic or incorrect answers), "?" (high frequency of "?" answers), and Mixed (no definite tendencies in the answers).

By means of a specially constructed apparatus, the 40 selected items were represented to 74 of the subjects who took the group tests, under conditions such that simultaneous measurements of psychogalvanic response and reaction time could be made, while recall was tested by the reproduction method immediately following the conclusion of the series of 40 items.

The testing of the first 20 of these subjects constituted experiment 1, regarded as more or less of an introductory nature. The testing of the last 50 subjects constituted experiment 2, from the results of which the main conclusions are drawn. This group

of 50 included 30 women and 20 men. Half of each sex group responded to the items arranged in one order, and the other half of each sex group responded to the items arranged in another order (belts 2n and 3n respectively). Also, half of each sex group gave the "Yes", "No" and "?" answers merely by pressing keys, while the other half spoke the answers in addition to pressing the keys (the vocal and non-vocal groups respectively). The subjects also indicated whether or not they considered each item difficult to answer.

Psychogalvanic responses were measured according to the technique described by Wechsler (24), and by means of an adaptation of the Wechsler Psychogalvanograph, with liquid electrodes. Reaction times and answers were taken on a specially constructed polygraph recorder.

Results. The more important findings and indications are recapitulated below.

(a). Results of the Group Testing.

1. The 125 items of the Bernreuter Personality Inventory are answered in an average of 12.97 minutes, compared with 15.25 minutes for the 125 items of the General Information Test. The difference, 2.28 minutes, is completely reliable (C.R. 7.35).

2. The "?" answer is utilized 6.96 times on the average in the Bernreuter Personality Inventory, compared with 23.32 times in the General Information Test. The difference, 16.36 is completely reliable (C.R. 10.22).

3. The Bernreuter scores (Bl-N scale) show a trend toward bimodality of distribution, compared with a fair degree of symmetry in the General Information scores distribution. The Bernreuter scores are nearly 4 times as variable as the General Information scores, and the sex difference is much larger and more reliable in the former.

4. The average number of obviously changed answers is 2.10 per subject for the Bernreuter Inventory, and 2.47 for the General Information Test. The difference, .37, has a C.R. of 1.94, or 97 chances out of 100 that it is a true difference greater than zero.

5. The Bernreuter scores are more variable than the General Information scores with respect to the times, the number of "?"

answers, and the number of changed answers.
(b). Psychogalvanic Response Results.

1. In experiment 1, the answers to personality items were accompanied by measurable deflections in 53.4% of the cases, compared with 47.7% for the achievement items. The mean personality deflection was 5.89, and the mean achievement deflection was 4.96, both in scale divisions corresponding approximately to millimeters on the Psychogalvanograph scale. The psychogalvanic results of experiment 1 indicated certain necessary changes in procedure which were employed in experiment 2.

2. All of the remaining statements concerning psychogalvanic responses refer to the results of experiment 2. Graphical comparison between the distributions of the deflections accompanying the responses to personality and achievement items show that there are fewer zero or very small deflections for the personality items, and more of the relatively large deflections. The reverse is true for the achievement items.

3. In the second half (Parts 3 and 4) of the stimulus series, after the positional effect has been largely eliminated or exhausted, the mean deflections following critical points tend to increase when the shift is from an achievement item to a personality item, and to decrease when the shift is from a personality item to an achievement item.

4. Individual subject average personality deflections exceed the individual subject average achievement deflections for the stimulus series as a whole and for each of the four Parts (successive groups of ten items each). This difference is least reliable for Parts 1 and 2 (C.R. .36) and most reliable for Part 3 (C.R. 2.23, indicating 99 out of 100 chances that the difference is greater than zero).

5. Individual subject averages either agree with the group tendency (larger personality deflections) or show excesses in favor of the achievement items which are smaller on the average than the excesses favoring the personality items.

6. Personality deflections, as a rule, are more variable than achievement deflections (measured by Pearson's Coefficient of Variation).

7. Individual subject mean deflections average 12.591 for the men and 7.175 for the women. The difference, 5.416, has a C.R. of 2.00 (or 98 chances out of 100 of being a real difference greater than zero). This sex difference is neither relatively larger nor markedly more reliable for the personality items than for the achievement items. The personality-achievement difference holds for both sexes,

but is neither relatively larger nor more markedly reliable for either sex.

8. Individual subject deflections average 12.727 for the group who vocalized the answers, and 5.966 for the group who did not vocalize the answers. The difference has a C.R. of 2.86 (99 chances out of 100 that it is real). The personality-achievement difference is relatively twice as large for the vocal group as for the non-vocal group, and is also slightly more reliable, but both the differences and the reliabilities are much smaller than for the vocalization difference itself. The vocalization difference is neither relatively larger nor more reliable for the personality items than for the achievement items, but has a reliability in each case which is slightly lower (C.R.'s of 2.59 and 2.72) than the reliability of the general vocalization difference.

(c). Reaction Time Results.

1. In experiment 1, the average personality reaction time was 6.48 seconds, and the average achievement reaction time was 6.62 seconds. The remaining statements regarding reaction times refer to experiment 2.

2. There is a tendency for large reaction times to be associated with achievement items more often than with personality items, while the short reaction times are more often associated with personality items.

3. Neither serial position nor length of item appears to be significant as a determiner of the length of the reaction times measured in this experiment.

4. Achievement times (i.e., reaction times) are longer, on the average, than personality reaction times for all items, for three of the four selected classes of items, for all subjects, for both separate-belt groups, for both vocalization groups, and for both sex groups.

5. The average of the individual subject mean personality reaction time is 7.506 seconds, and of the individual subject mean achievement reaction times, 7.857 seconds. The difference, .351 seconds, has a C.R. of 1.95 (97 chances out of 100 that it is a true difference greater than zero).

6. The average of the individual subject mean reaction times for the men is 8.629, and for the women, 7.026 seconds. The difference, 1.603 seconds, has a C.R. of 2.29 (99 out of 100 chances that it is a real difference greater than zero). The sex difference is neither relatively greater nor more reliable for either type of item; nor is the personality-achievement difference relatively greater or more reliable for either sex (except as the larger number of women makes the

difference appear more reliable for them).

7. The individual subject mean reaction times for the vocal group average 7.484 seconds, and for the non-vocal group, 7.878 seconds. The difference, .394 seconds, has a C.R. of .56 (71 chances out of 100 that it is a true difference greater than zero). The personality-achievement difference is relatively larger for the vocal group than for the non-vocal group but is not markedly more reliable for either group although the reliabilities for both groups are somewhat higher (C.R.'s .90 and .83) than the reliability of the vocalization difference itself.

8. Reaction time is not related significantly to the psychogalvanic responses.

9. There are opposite type tendencies in the reaction times for the four selected classes of items, especially in the Favorable and Unfavorable classes.

10. Relative variabilities are nearly equal for the two types of items, but are greater for the women and the vocal group.

(d). Memory Value Results.

1. In experiment 1, the average number of personality items recalled was 7.65, and achievement items, 6.35, while 14 of the 20 subjects agreed with the group tendency, 3 recalled the same number of items of both types, and only 3 recalled a larger number of achievement items. The remaining statements regarding memory values apply to the results of experiment 2.

2. For all subjects of experiment 2, the average number of personality items recalled was 7.160, and the average number of achievement items recalled was 6.340. Personality memory scores are larger than achievement memory scores for all subjects, for the women, for both vocalization groups, and for the Unfavorable and "?" classes of items.

3. The personality-achievement difference for all subjects is .820, with a C.R. of 2.64 (99 out of 100 chances that it is a true difference greater than zero).

4. The total memory score for the women is 14.233, and for the men, 12.400. The difference, 1.833, has a C.R. of 1.74 (96 out of 100 chances that it is a true difference greater than zero). The sex difference holds true for both types of items and for both vocalization groups.

5. Personality and achievement memory scores are equal for the men. For the women, the scores are respectively 7.800 and 6.433. The difference of 1.367 has a C.R. of 3.33 which indicates complete reliability.

6. Achievement memory scores are nearly equal for the men and the women. The

personality memory scores for the two sexes are respectively 6.200 and 7.800. The difference, 1.600, has a C.R. of 2.85 (99 chances out of 100 that it is a true difference greater than zero).

7. The difference in total memory scores between the vocal and non-vocal groups is negligible, but the personality-achievement difference is nearly twice as large for the vocal group, and also more reliable (C.R.'s for the two groups are 2.00 and 1.15 respectively).

8. Serial position was not found to be a decisive factor influencing the memory scores.

9. Memory scores are highest for the Unfavorable personality class of items.

10. Memory scores do not appear to be related significantly to either the psychogalvanic responses or the reaction times. The lack of interrelationships among the three principal measures used in this investigation indicates that different groups of subjects are responsible for the three main tendencies found, although nearly all subjects agree with at least one of the group tendencies and 14 subjects agree with all three of the group tendencies.

11. Judgments of difficult are associated more often with achievement items (C.R. 1.42, or 92 out of 100 chances of being a real difference greater than zero). There appear to be type differences in the classes of items judged difficult.

12. Changed answers in experiment 2 were too few for any study of type differences, but the general tendencies were for greatly increased deflections and reaction times.

General Summary of Results. The measures taken in this study have been found to be influenced in varying degrees by several factors (serial position, sex of subject, and vocalization of answers) in addition to the type of item (personality or achievement). Differences due to the latter were the chief concern, and the data seem to justify the following tentative conclusions:

1. For the pooled results from various groups of subjects, the psychogalvanic responses accompanying the answers to personality items are slightly larger (amplitude of mean deflection) and more variable than the psychogalvanic responses accompanying the answers to achievement items.

2. For the pooled results from various groups of subjects, the reaction times associated with the answers to achievement items are slightly longer than the reaction times associated with the answers to personality items. The relative variabilities are about equal.

3. The women and the group as a whole recall more personality items than achievement items. Achievement memory scores tend to be more variable than personality memory scores.

4. Responses to personality items are characterized by fewer uses of the "?" answer, by fewer judgments of "Difficult", and by fewer changes of answer, as compared with the responses to achievement items.

Interpretation of Results. Since a careful search through the literature failed to reveal any studies making the sort of comparison attempted herein, the purpose of the present investigation was deemed to be largely exploratory, placing major emphasis on the "whether" rather than on the "what" with regard to differences between the two types of test items.

It is believed that the data justify an affirmative answer to the question of whether the responses to the two types of items affect differently the measures which were used as differentiating agents, although the differences are admittedly small, and it is further believed that the conclusions presented in the preceding General Summary of Results indicate the probable directions of the differences between the two types of items. However, in line with the point of view taken in the first Chapter, it is recognized that the findings are tentative, and may be applied strictly only to the specific materials used in this study, less certainly to true-false achievement and yes-no adjustment questions as a whole, and only by the wildest stretch of imagination to achievement questions and personality questions in general. The study has dealt with a very limited aspect of an extremely wide and complicated situation, and strongly positive statements will be justified only after much more research along the general lines laid down herein.

What, then, is the contribution made by the present investigation? The contribution has a threefold aspect. In the first place, it has been demonstrated that the important issue concerning the validity of "objective" techniques in personality testing (an issue which up to the present time has been treated chiefly through controversial theorizing) may be approached experimentally by a direct comparison of responses to personality and achievement items. This is a contribution to technique. In the second place, it has been demonstrated that, in terms of certain commonly used measures (and others), there probably are definite although small differences in the subjective conditions influencing responses to the two types of test

items; and the results have indicated the probable directions in which the subjective conditions influence the measures. This is at least a tentative contribution to existing knowledge. In the third place, the results raise certain problems well worthy of being studied further, concerning which more will be said presently.

The discussions have purposely refrained from any extensive attempt to analyze the qualitative nature of the subjective factors which were predicated as the responsible agents underlying the differences in deflections, reaction times, memory values, use of the "?" answer, judgments of "Difficult", and changed answers. However, logical completeness demands that some attention, at least, be given to this aspect of the situation. Let the psychogalvanic response be considered first. Students of this phenomenon have differed widely in their interpretations of its psychological correlates. The emotion theory has had no lack of advocates--e.g., Smith (20) and Wechsler (24)--while a few have held to the conation theory, notably Aveling (4). In his most recent extensive review of the whole problem, Landis (25) goes to the other extreme and chooses to regard the psychogalvanic response as purely physiological in significance. Probably the most reasonable and conservative viewpoint is reflected by Sears (19), who holds that the psychogalvanic response in itself indicates only hyper-activity of the autonomic nervous system, sympathetic division, and that the psychological correlates will vary with different situations, although there seems to be a common substratum which may be best described as "feeling-tension." If this point of view is accepted, the findings of this study would indicate a slightly greater expenditure of nervous energy in connection with the answering of personality items. However, the difference between the two types in respect to psychogalvanic responses is quite small (both absolutely and relative to the sex and vocalization differences), and it is subject to wide individual variations both in direction and magnitude. It may well be doubted, therefore, whether any subjective differences underlying the psychogalvanic difference would be sufficient to justify the conclusion that the personality testing situation is loaded with any more disturbing elements than is the achievement testing situation. Indeed, as was pointed out in Chapter III, the fact that there are fewer "?" answers and fewer changes of answer on the personality test would seem to indicate that the increased feeling-tension, if real, serves to facilitate

rather than to hinder efficient test performance.

On the other hand, the evidence offered by the memory scores points in part at least to the opposite conclusion. There has been a great deal of research regarding the supposed relationships between memory and affectivity. Meltzer (15), in his 1930 review, concluded that the value of most of the evidence at that time was practically nil, and the findings of later studies are conflicting. It should be noted, however, that the disagreement does not concern whether memory is conditioned by some subjective states (the only claim that has been made for its use in this study). Rather, the argument concerns whether these subjective states are primarily affective in nature. The requirements of this study do not make it necessary to take sides in the matter. The personality-achievement difference in memory scores is the most reliable of all of those found in experiment 2, indicating that the subjective states underlying the responses--whatever those states may be--are more sharply differentiated by the memory scores than by the deflections or the reaction times. Also, the unusually high memory value of the Unfavorable personality items might be interpreted as being due to conflict of a disturbing nature.

The reaction time data, like the psychogalvanic response data, are not particularly conclusive. The very long or the very short reaction times do not accompany either type of item exclusively, and the only coincidence between long reaction times and large deflections comes in connection with the changed answers, where it is the same for both types. However, the shorter personality reaction time adds another element to the picture of apparently greater efficiency in the personality testing situation.

On the whole, therefore, such differences as have been found between the reactions to personality and achievement test items are not sufficiently large or sufficiently consistent from subject to subject to justify the assumption that the subjective conditions affecting the responses to either type of item are markedly different from those affecting the responses to the other type of item. Accordingly, insofar as it is permissible to generalize on the basis of the limited data, the theoretical criticisms of the personality test (or at least that aspect in which the study were primarily interested) do not gain much support from the findings of this study.

Problems for Further Research. Every finding suggested by the results of this study is sufficiently tentative to offer an

opportunity for further investigation; only a few can be considered here.

One of the most interesting issues is raised by the relatively large difference in mean deflections for the vocal and non-vocal groups. Oral answers have been employed almost universally in experiments combining the psychogalvanic and free-association techniques, also in the studies of the effects of mental work on the psychogalvanic response. In fact, vocal answers seem to have been taken for granted, and there is little or no recognition of vocalization as a possibly important factor in the production of the deflections. The data of this study suggest that the deflections measured in such experiments may have been due in large part to the vocalization factor rather than to any other factors. It would be interesting to have some of these experiments repeated with some silent method (such as the use of keys) of giving the answers, although it is difficult to see just how this could be accomplished in the case of the free-association experiment. In the present study, the "vocal" and "non-vocal" deflections were measured on different groups. The same set-up ought to be studied more extensively, with the addition of a third group in which the same subjects gave part of their answers orally and part of them silently.

The memory value results offer several possibilities for further investigation. The apparent sex difference is worth more intensive study, especially in relation to such factors as intelligence. The memory values ought also to be tested by delayed recall as well as by immediate recall.

The study has entirely neglected the rich possibilities of a careful study of the introspective data which might be obtained from the subjects in such an experiment as this, perhaps by means of a special questionnaire to be filled out by the subjects after the individual testing.

The basic technique of this study should be applied in experiments comparing the results for other types of items, such as the multiple choice. It would also be worth while to repeat the whole experiment with several large groups in which the sex and vocalization factors were held constant.

Finally, no claim has been made or inferred that the measuring instruments employed herein are the only ones--or perhaps even the most important ones--which might and should be brought to bear on the general problem of subjective differences in the two testing situations. For example, some other physiological measure, such as blood pressure,

might serve the purpose even better. The choice of measures to be used in this study was a matter of convenience and availability rather than of necessity. The writer is convinced, however, that the fundamental idea of

this investigation--direct comparison of reactions to the two types of test items--is worthy of much wider application than it has been possible to make in the present treatment.

BIBLIOGRAPHY

This Bibliography includes only those references to which some reference has been made in the body of the dissertation.

1. Abel, T. M. "Attitudes and the Galvanic Skin Reflex", Journal of Experimental Psychology, XIII (Feb., 1930) 47-60.
2. Allport, G. W., and Vernon, P. E. "The Field of Personality", Psychological Bulletin, XXVII (Dec., 1930), 677-730.
3. American Educational Research Association. "Tests of Personality and Character", Review of Educational Research, II (June, 1932), 270 pp.
4. Aveling, F. Directing Mental Energy. New York: Doran, 1927.
5. Colvin, S. S. "Principles Underlying the Construction and Use of Intelligence Tests", Chapter II of Intelligence Tests and Their Use. Twenty-first Yearbook of the National Society for the study of Education, Part I. Bloomington, Illinois: Public School Publishing Co., 1923. Pp. 275.
6. Crosland, H. R., and Beck, L. F. Objective Measurements of Emotion. University of Oregon Publication, Psychology Series, I (Dec., 1931). Pp. 202.
7. Darrow, C. W. "Electrical and Circulatory Responses to Brief Sensory and Ideational Stimuli", Journal of Experimental Psychology, XII (Aug., 1929), 267-300.
8. Garrett, H. E. Statistics in Psychology and Education. New York: Longmans, Green & Co., 1926. Pp. 317.
9. Gilliland, A. R. "Problems of Personality." Journal of Abnormal and Social Psychology, XXIII (Oct.-Dec., 1928), 369-378.
10. Hathaway, S. R. "A Comparative Study of Psychogalvanic and Association Measures." Journal of Applied Psychology, XIII (1929), 632-646.
11. Hildreth, G. H. A Bibliography of Mental Tests and Rating Scales. New York: Psychological Corporation, 1933. P. 242.
12. Jones, H. E. "Emotional Factors in Learning", Journal of General Psychology, II (April-July, 1929), 263-270.
13. Lynch, C. A. "The Memory Values of Certain Alleged Emotionally Toned Words", Journal of Experimental Psychology, XV (June, 1932), 298-315.
14. Manson, G. E. A Bibliography of the Analysis and Measurement of Human Personality up to 1926. Reprint and Circular Series, National Research Council, 1926, No. 72.
15. Meltzer, H. "The Present Status of Experimental Studies on the Relationship of Feeling to Memory", Psychological Review, XXXVII (March, 1930), 124-139.
16. National Education Association, Department of Superintendence. Tenth Yearbook, Character Education. Washington, D. C.: Department of Superintendence, 1932. Pp. 535.
17. Raubenheimer, A. S. "An Experimental Study of Some Behavior Traits of the Potentially Delinquent Boy", Psychological Monographs, XXXIV (1925, No. 5), Pp. 106.
18. Rugg, H. O. Statistical Methods Applied to Education. Boston: Houghton Mifflin Co., 1917. Pp. 410.
19. Sears, R. "Psychogalvanic Responses in Arithmetical Work: Effects of Experimental Changes in Addition", Archives of Psychology, XXIV (March, 1933), Pp. 62.
20. Smith, Whately. The Measurement of Emotion. New York: Harcourt, Brace & Co., 1922. Pp. 184.
21. Symonds, P. M. Diagnosing Personality and Conduct. New York: Century Co., 1931. Pp. 602.
22. Thomas, W. I., and Thomas, D. S. The Child in America. New York: Alfred A. Knopf, 1928. Pp. 583.
23. Watson, Goodwin. "Character and Personality Tests." Psychological Bulletin, XXX (July, 1933), 467-487.
24. Wechsler, D. "The Measurement of Emotional Reactions", Archives of Psychology, XII (1925), No. 76. Pp. 181.
25. Landis, Carney. "Electrical Phenomena of the Skin", Psychological Bulletin, XXIX (Dec., 1932), 693-752.

Note: Since this investigation was interested in the pgr, reaction time, and memory values only in an instrumental sense, it was not deemed worth while to include bibliographies dealing particularly with those measures. Such bibliographies may be found, however, in some of the above references, as follows: pgr, 3, 6, 21, 26; reaction time, 6, 21; memory values, 15.

APPENDIX A

Order of Items in Belts 1, 4, and 5, also Belts 2n and 3n

Order of Items in Belt 1

Part 1.

Do you usually try to avoid bossy or dictatorial people?
Do you usually work things out for yourself rather than get someone to show you?
Are people sometimes successful in taking advantage of you?
Do people ever come to you for advice?
Do you like to get many views from others before making an important decision?
Are you considered to be critical of other people?
Would you "have it out" with a person who spread untrue rumors about you?
Do you day-dream frequently?
When you are in low spirits do you try to find someone to cheer you up?
Does admiration gratify you more than achievement?
Do you feel that marriage is essential to your present or future happiness?
Do you prefer a play to a dance?
Do you try to treat a domineering person the same as he treats you?
Do you ever heckle or question a public speaker?
Have you ever crossed the street to avoid meeting some person?
Do you prefer to be alone at times of emotional stress?
Are you easily discouraged when the opinions of others differ from your own?
Do you very much mind taking back articles you have purchased at stores?
If you came late to a meeting would you rather stand than take a front seat?
Do you try to get your own way even if you have to fight for it?

Part 2.

Was Dickens interested in the condition of the poor?
Is the anteater an edentate animal?
Is Valley Forge situated in New Jersey near Trenton?
Are proteins a necessary food element?
Is Wolfgang Köhler one of the better-known psychoanalysts?

Was President Garfield shot by the assassin Czolgosz?
Is Bogota the capital of the Republic of Venezuela?
Are spiders classified in the same phylum with crabs?
Are there any nudists among the Dukhobors?
Was St. Paul one of the original twelve Disciples?
Does USSR mean "United Soviet States of Russia"?
Was "The Gold Bug" a tale of the California gold-rush?
Is the mesencephalon a neurological structure?
Does a pastoral usually deal with country or farm life?
Is Popocatepetl in the same country as Chichen-Itza?
Can tobacco be cultivated successfully in northern States?
Is the word "supercede" spelled correctly?
Is the base of a triangle always the longest side?
Are postulates proved in geometrical reasoning?
Does union labor favor the "closed shop" policy?

Order of Items in Belt 4

Part 1 is the same as Part 2 of Belt 1, in reverse order.
Part 2 is the same as Part 1 of Belt 1, in reverse order.

Order of Items in Belt 5

Part 1.

Do you like to get many views from others before making an important decision?
Are there any nudists among the Dukhobors?
Are you considered to be critical of other people?
Are spiders classified in the same phylum with crabs?
Is Bogota the capital of the Republic of Venezuela?
Do people ever come to you for advice?
Would you "have it out" with a person who spread untrue rumors about you?
Do you try to get your own way even if you

have to fight for it?
 Was President Garfield shot by the assassin Czolgosz?
 Do you day-dream frequently?
 Was Dickens interested in the condition of the poor?
 Was St. Paul one of the original twelve Disciples?
 Does a pastoral usually deal with country or farm life?
 When you are in low spirits do you try to find someone to cheer you up?
 Does union labor favor the "closed shop" policy?
 Does admiration gratify you more than achievement?
 Is the word "supercede" spelled correctly?
 Is the anteater an edentate animal?
 Do you feel that marriage is essential to your present or future happiness?
 Are people sometimes successful in taking advantage of you?

Order of Items in Belt 5, Part 2

Are proteins a necessary food element?
 Do you usually work things out for yourself rather than get someone to show you?
 Is Wolfgang Köhler one of the better-known psychoanalysts?
 Do you try to treat a domineering person the same as he treats you?
 Do you ever heckle or question a public speaker?
 Is Valley Forge situated in New Jersey near Trenton?
 Does USSR mean "United Soviet States of Russia"?
 Are postulates proved in geometrical reasoning?
 Have you ever crossed the street to avoid meeting some person?
 Is Popocatepetl in the same country as Chichen-Itza?
 Do you prefer to be alone at times of emotional stress?
 Are you easily discouraged when the opinions of others differ from your own?
 If you came late to a meeting would you rather stand than take a front seat?
 Is the base of a triangle always the longest side?
 Do you prefer a play to a dance?
 Can tobacco be cultivated successfully in northern States?
 Do you usually try to avoid bossy or dictatorial people?
 Do you very much mind taking back articles you have purchased at stores?
 Was "The Gold Bug" a tale of the California gold-rush?

Is the mesencephalon a neurological structure?

Order of Items in Belt 2n

Part 1.

Was dickens interested in the condition of the poor?
 Is the anteater an edentate animal?
 Is Valley Forge situated in New Jersey near Trenton?
 Are proteins a necessary food element?
 Is Wolfgang Köhler one of the better-known psychoanalysts?
 Was President Garfield shot by the assassin Czolgosz?
 Is Bogota the capital of the Republic of Venezuela?
 Are spiders classified in the same phylum with crabs?
 Are there any nudists among the Dukhobors?
 Was St. Paul one of the original twelve Disciples?

Part 2.

Do you day-dream frequently?
 Does admiration gratify you more than achievement?
 Would you "have it out" with a person who spread untrue rumors about you?
 Do you feel that marriage is essential to your present or future happiness?
 Do you prefer to be alone at times of emotional stress?
 Do you try to get your own way even if you have to fight for it?
 Are people sometimes successful in taking advantage of you?
 Are you considered to be critical of other people?
 Have you ever crossed the street to avoid meeting some person?
 Do you usually try to avoid dictatorial or bossy people?

Part 3.

Does USSR mean "United Soviet States of Russia"?
 Do people ever come to you for advice?
 Is the word "supercede" spelled correctly?
 Are you easily discouraged when the opinions of others differ from your own?
 Is the mesencephalon a neurological structure?
 Is the base of a triangle always the longest side?
 If you came late to a meeting would you rather stand than take a front seat?

Do you prefer a play to a dance?
Does union labor favor the "closed shop" policy?
Do you very much mind taking back articles you have purchased at stores?

Part 4.

Do you ever heckle or question a public speaker?
Do you try to treat a domineering person the same as he treats you?
Do you usually work things out for yourself rather than get someone to show you?
When you are in low spirits do you try to find someone to cheer you up?
Do you like to get many views from others before making an important decision?
Was "The Gold Bug" a tale of the California gold-rush?
Does a pastoral usually deal with country or farm life?
Is Popocatepetl in the same country as Chichen-Itza?
Can tobacco be cultivated successfully in northern States?
Are postulates proved in geometrical reasoning?

Order of Items in Belt 3n

Part 1.

Do people ever come to you for advice?
Are you easily discouraged when the opinions of others differ from your own?
If you came late to a meeting would you rather stand than take a front seat?
Do you prefer a play to a dance?
Do you very much mind taking back articles you have purchased at stores?
Do you ever heckle or question a public speaker?
Do you try to treat a domineering person the same as he treats you?
Do you usually work things out for yourself rather than get someone to show you?
When you are in low spirits do you try to find someone to cheer you up?
Do you like to get many views from others before making an important decision?

Part 2.

Are postulates proved in geometrical reasoning?
Can tobacco be cultivated successfully in northern States?

Is Popocatepetl in the same country as Chichen-Itza?
Does a pastoral usually deal with country or farm life?
Was "The Gold Bug" a tale of the California gold-rush?
Does union labor favor the "closed shop" policy?
Is the base of a triangle always the longest side?
Is the mesencephalon a neurological structure?
Is the word "supercede" spelled correctly?
Does USSR mean "United Soviet States of Russia"?

Part 3.

Do you day-dream frequently?
Was Dickens interested in the condition of the poor?
Does admiration gratify you more than achievement?
Is the anteater an edentate animal?
Would you "have it out" with a person who spread untrue rumors about you?
Do you feel that marriage is essential to your present or future happiness?
Is Valley Forge situated in New Jersey near Trenton?
Are proteins a necessary food element?
Do you prefer to be alone at times of emotional stress?
Is Wolfgang Köhler one of the better-known psychoanalysts?

Part 4.

Was President Garfield shot by the assassin Czolgosz?
Is Bogota the capital of the Republic of Venezuela?
Are spiders classified in the same phylum with crabs?
Was St. Paul one of the original twelve Disciples?
Are there any nudists among the Dukhobors?
Do you usually try to avoid dictatorial or "bossy" people?
Have you ever crossed the street to avoid meeting someone?
Are you considered to be critical of other people?
Are people sometimes successful in taking advantage of you?
Do you try to get your own way even if you have to fight for it?

APPENDIX B

Tables

Table I-A
"YES", "NO", "?" AND OMITTED ANSWERS GIVEN BY 86 COLLEGE STUDENTS
TO THE ITEMS IN THE BERNREUTER PERSONALITY INVENTORY

ITEM	YES	NO	?	OM	ITEM	YES	NO	?	OM
1.	41	42	2	1	47.	11	70	5	0
2.	54	31	1	0	48.	25	58	3	0
3.	81	4	1	0	49.	35	49	2	0
4.	67	18	1	0	50.	48	37	1	0
5.	57	23	4	2	51.	30	54	2	0
6.	62	23	1	0	52.	50	33	3	0
7.	9	72	4	1	53.	43	36	7	0
8.	37	48	1	0	54.	24	60	1	1
9.	12	74	0	0	55.	61	20	5	0
10.	6	78	2	0	56.	72	9	5	0
11.	39	41	5	1	57.	25	57	3	1
12.	21	58	6	1	58.	31	52	3	0
13.	19	63	4	0	59.	47	34	4	1
14.	28	56	1	1	60.	35	51	0	0
15.	55	30	1	0	61.	41	39	6	0
16.	32	53	1	0	62.	60	20	6	0
17.	45	35	4	2	63.	17	65	3	1
18.	32	52	1	1	64.	56	29	1	0
19.	5	78	3	0	65.	26	57	3	0
20.	36	50	0	0	66.	25	57	3	1
21.	49	36	1	0	67.	82	1	3	0
22.	30	47	7	2	68.	9	68	9	0
23.	28	55	3	0	69.	25	57	4	0
24.	24	57	5	0	70.	44	38	3	1
25.	57	24	5	0	71.	45	32	9	0
26.	20	64	2	0	72.	33	49	4	0
27.	22	61	3	0	73.	23	56	7	0
28.	32	44	10	0	74.	26	54	5	1
29.	11	73	2	0	75.	42	37	7	0
30.	42	43	1	0	76.	26	58	2	0
31.	47	35	4	0	77.	40	35	11	0
32.	18	60	8	0	78.	45	40	1	0
33.	36	41	8	1	79.	41	38	7	0
34.	28	49	8	1	80.	26	53	7	0
35.	29	52	4	1	81.	8	74	4	0
36.	48	38	0	0	82.	62	16	8	0
37.	66	17	3	0	83.	61	22	3	0
38.	49	30	7	0	84.	26	56	4	0
39.	35	50	1	0	85.	67	14	5	0
40.	52	33	1	0	86.	53	24	9	0
41.	58	14	14	0	87.	65	13	8	0
42.	43	37	6	0	88.	40	42	4	0
43.	46	34	6	0	89.	69	17	0	0
44.	23	57	6	0	90.	41	41	4	0
45.	30	54	2	0	91.	27	57	2	0
46.	36	41	9	0	92.	63	21	2	0

Table I-A (Continued).

ITEM	YES	NO	?	OM	ITEM	YES	NO	?	OM
93.	18	64	4	0	109.	20	59	7	0
94.	56	23	6	1	110.	55	28	2	1
95.	66	10	10	0	111.	64	22	0	0
96.	10	70	6	0	112.	42	43	1	0
97.	48	31	7	0	113.	23	53	10	0
98.	40	35	11	0	114.	16	66	4	0
99.	13	67	5	1	115.	24	58	4	0
100.	65	16	5	0	116.	29	46	11	0
101.	51	29	6	0	117.	44	33	7	2
102.	61	18	7	0	118.	24	52	10	0
103.	24	59	3	0	119.	49	33	4	0
104.	32	50	4	0	120.	49	25	12	0
105.	63	18	5	0	121.	64	15	7	0
106.	32	52	2	0	122.	66	14	6	0
107.	66	18	2	0	123.	26	56	4	0
108.	55	27	4	0	124.	12	69	5	0
					125.	20	61	5	0

Table II-A
 "YES", "NO", "?" AND OMITTED ANSWERS GIVEN BY 86 COLLEGE STUDENTS
 TO THE ITEMS IN THE GENERAL INFORMATION TEST

ITEM	YES	NO	?	OM	ITEM	YES	NO	?	OM
1.	55	26	4	1	30.	47	30	8	1
2.	23	48	14	1	31.	8	29	48	1
3.	2	80	4	0	32.	22	29	34	1
4.	19	25	39	3	33.	25	44	17	0
5.	43	9	34	0	34.	14	14	55	3
6.	6	77	3	0	35.	10	11	65	0
7.	38	39	8	1	36.	21	4	58	3
8.	69	14	3	0	37.	67	14	5	0
9.	84	0	1	1	38.	15	58	12	1
10.	75	5	5	1	39.	44	24	17	1
11.	3	77	6	0	40.	42	4	40	0
12.	6	43	36	1	41.	4	55	27	0
13.	82	2	1	1	42.	17	65	3	1
14.	14	53	18	1	43.	51	11	22	2
15.	47	25	11	3	44.	6	70	10	0
16.	67	14	5	0	45.	47	14	25	0
17.	5	81	0	0	46.	61	19	6	0
18.	33	53	0	0	47.	17	61	8	0
19.	64	19	2	1	48.	20	39	27	0
20.	12	33	40	1	49.	13	60	13	0
21.	24	46	15	1	50.	68	15	3	0
22.	41	21	24	0	51.	28	41	17	0
23.	44	13	28	1	52.	32	23	31	0
24.	6	71	8	1	53.	48	30	8	0
25.	1	85	0	0	54.	18	12	55	1
26.	48	14	24	0	55.	33	37	16	0
27.	65	16	5	0	56.	49	31	6	0
28.	8	43	34	1	57.	58	22	5	1
29.	30	55	1	0	58.	0	86	0	0

Table II-A (Continued).

ITEM	YES	NO	?	OM	ITEM	YES	NO	?	OM
59.	86	0	0	0	93.	43	40	2	1
60.	28	50	7	1	94.	20	9	57	0
61.	24	56	6	0	95.	51	26	9	0
62.	68	12	6	0	96.	58	26	1	1
63.	41	39	4	2	97.	19	64	3	0
64.	1	85	0	0	98.	1	76	9	0
65.	33	26	27	0	99.	47	15	24	0
66.	69	9	8	0	100.	9	76	1	0
67.	11	42	33	0	101.	63	20	3	0
68.	51	24	11	0	102.	6	79	1	0
69.	6	77	3	0	103.	59	15	12	0
70.	30	47	9	0	104.	43	31	12	0
71.	73	9	3	1	105.	11	32	42	1
72.	45	10	30	1	106.	16	56	14	0
73.	71	2	13	0	107.	73	2	10	1
74.	73	9	3	1	108.	43	31	12	0
75.	9	67	10	0	109.	51	23	12	0
76.	16	36	34	0	110.	2	80	4	0
77.	27	19	40	0	111.	55	21	10	0
78.	53	3	30	0	112.	23	31	32	0
79.	68	5	13	0	113.	55	12	19	0
80.	11	62	13	0	114.	27	51	8	0
81.	12	55	19	0	115.	47	23	15	1
82.	21	33	32	0	116.	45	35	6	0
83.	2	80	3	1	117.	62	5	19	0
84.	37	29	20	0	118.	65	11	10	0
85.	47	24	14	1	119.	60	15	10	1
86.	29	39	18	0	120.	74	1	11	0
87.	33	25	28	0	121.	65	10	11	0
88.	41	12	33	0	122.	80	3	3	0
89.	29	40	17	0	123.	59	11	15	1
90.	13	48	25	0	124.	32	17	37	0
91.	40	29	15	2	125.	67	5	14	0
92.	27	46	12	1					

Table III-A
 "YES", "NO", "?" AND OMITTED ANSWERS, ALSO THE JUDGMENTS
 OF "DIFFICULT" GIVEN BY SEVENTY SUBJECTS TO THE SELECTED
 TEST ITEMS USED IN EXPERIMENTS 1 AND 2 (PERSONALITY)

Item No.	YES	NO	?	OMITTED	DIFFICULT	Item No.	YES	NO	?	OMITTED	DIFFICULT
2	41	26*	2*	1	7	78	37*	24	9*	0	8
3	60*	3	7	0	7	81	7	55*	8	0	13
4	55	15*	0	0	4	88	29	35*	6*	0	4
10	9	56*	5*	0	16	89	58	10	2*	0	4
11	25*	37	8*	0	11	90	30	29	10*	1	9
29	11	57*	2*	0	3	95	48*	11	11	0	11
30	33	34*	3	0	6	98	30	28*	12*	0	12
37	50	17*	3*	0	10	100	48	18*	4*	0	10
67	67*	1	2*	0	4	116	25	30*	15	0	8
77	44	19*	7*	0	7	120	40*	19	10	1	8

*Indicates favorable (normal) answer, or positive weight on the BI-N scale.

Table IV-A
 "YES", "NO", "?" AND OMITTED ANSWERS, ALSO JUDGMENTS OF "DIFFICULT"
 GIVEN SEVENTY SUBJECTS TO THE SELECTED TEST ITEMS USED IN EXPER-
 IMENTS 1 AND 2 (ACHIEVEMENT)

Item No.	YES	NO	?	OMITTED	DIFFICULT	Item No.	YES	NO	?	OMITTED	DIFFICULT
1	46	17*	7	0	10	59	68*	1	0	1	2
9	69*	0	1	0	3	65	31*	18	21	0	15
17	2	66*	2	0	3	82	13	29*	28	0	17
29	25*	39	4	2	8	94	12*	7	51	0	20
32	17	14*	37	2	19	95	39	23*	7	1	4
34	9*	6	55	0	20	101	54	15*	1	0	9
35	5	6*	58	1	23	109	40	16*	14	0	5
36	11*	5	54	0	21	110	4	62*	3	1	1
48	21*	26	23	0	16	112	13	27*	29	1	17
54	17	8*	44	1	16	122	62*	3	4	1	2

*Indicates favorable (correct) answer.

Table V-A
 INDIVIDUAL SUBJECT RECORDS IN EXPERIMENT 2: MEAN PERSONALITY AND
 ACHIEVEMENT DEFLECTIONS, MEAN PERSONALITY AND ACHIEVEMENT REAC-
 TION TIMES, AND PERSONALITY AND ACHIEVEMENT MEMORY SCORES

Subject	Mean Deflections		Mean Reaction Times		Memory Scores	
	Pers.	Achieve.	Pers.	Achieve.	Pers.	Achieve.
24F2nV*	15.95	31.85	15.67	17.53	11	7
25M2nV	13.60	13.20	10.92	12.22	6	7
26F2nV	10.80	11.40	5.72	7.65	5	5
27M2nV	11.20	15.85	13.05	14.12	5	7
28F2nV	1.10	4.20	4.90	4.87	7	8
29M2nV	35.10	32.45	11.20	10.92	7	7
30M2nV	6.50	8.65	7.80	6.47	5	6
31F3nV	9.35	2.05	6.45	5.6	5	5
32F3nV	1.70	1.35	4.75	6.97	6	2
33M3nV	28.80	24.20	7.87	7.77	6	3
34M3nV	20.80	12.20	7.97	9.10	6	8
35F3nV	6.45	1.75	4.47	4.65	7	4
36F3nV	12.50	12.20	5.40	4.95	10	5
37M3nV	57.10	32.35	5.02	3.65	7	9
38M3nV	8.05	14.60	6.90	8.17	11	7
39F3nV	2.31	2.41	8.21	9.97	10	6
40F2n	.65	3.50	5.52	4.30	9	13
41F2nV	5.25	4.00	6.95	7.95	8	6
42F3n	23.20	5.55	7.55	5.90	7	5
43F3nV	16.35	9.00	5.70	5.85	13	6
44F2n	0.00	0.00	9.12	11.42	5	6
45F2nV	4.10	4.78	4.05	4.00	7	8
46M3n	8.55	7.01	6.97	8.72	5	4
47F3n	8.40	6.30	10.55	9.15	4	7
48F3n	2.00	1.45	4.25	4.30	4	5
49M3n	4.20	.85	11.62	12.07	7	4
50F3n	4.45	3.50	6.45	7.57	7	5

*The first two numbers indicate the order of appearance of the subjects. The letter M indicates that the subject was a male. F indicates female. 2n and 3n indicate the belt used. V means that the answers were given vocally.

Table V-A (Continued).

Subject	Mean Deflections		Mean Reaction Times		Memory Scores	
	Pers.	Achieve.	Pers.	Achieve.	Pers.	Achieve.
51M3n	5.37	4.45	10.87	9.00	6	8
52M3n	2.25	1.40	8.25	9.57	4	3
53F2n	18.1	17.05	12.40	7.95	9	9
54F2n	2.20	2.15	6.12	8.57	9	8
56F2n	.85	2.00	9.02	10.80	7	7
57F2n	10.55	10.25	7.12	6.95	11	8
58M2n	5.35	8.00	4.62	5.05	7	6
59F2n	15.70	21.55	6.52	6.75	4	5
60M2nV	20.65	18.90	8.15	7.27	4	8
61F2n	.10	1.20	6.81	5.80	10	8
62F3n	4.42	2.42	5.92	6.65	8	4
63F3nV	26.95	16.30	4.05	4.52	11	9
64F2n	1.60	2.35	7.77	9.40	9	11
65F2nV	2.00	2.05	6.50	5.95	5	3
66M3n	3.95	5.65	7.40	9.10	8	9
67F3nV	4.80	1.80	6.95	5.70	10	11
68M2n	14.25	10.70	6.77	7.27	5	4
69M2nV	4.75	9.85	4.00	5.30	7	5
70F3n	7.35	6.90	6.42	7.47	5	3
71F3nV	10.55	10.10	4.85	6.57	8	7
72M2n	2.00	4.60	9.35	9.90	4	6
73M2nV	9.25	12.25	11.57	12.50	5	3
74M3n	3.15	1.75	8.82	8.30	9	10

Table VI-A

INDIVIDUAL SUBJECT RECORDS IN EXPERIMENT 1: MEAN PERSONALITY AND ACHIEVEMENT DEFLECTIONS, MEAN PERSONALITY AND ACHIEVEMENT REACTION TIMES, AND PERSONALITY AND ACHIEVEMENT MEMORY SCORES

Subject	Mean Deflections		Mean Reaction Times		Memory Scores	
	Pers.	Achieve.	Pers.	Achieve.	Pers.	Achieve.
1F1*	.45	.45	5.53	7.10	6	6
2F4	2.15	2.45	5.62	5.90	9	9
3F5	7.65	2.55	5.26	5.00	5	2
4F1V	1.95	1.35	7.15	6.15	12	7
5F4V	1.20	.50	7.35	7.20	6	7
6F5V	2.70	4.40	4.70	4.57	9	7
7M1	1.50	.75	4.97	6.82	5	3
8F4	5.60	6.10	5.02	6.95	10	7
9F5	.52	.10	7.63	9.85	6	5
10F1V	9.25	1.73	7.32	6.18	7	6
11M4V	15.30	11.95	6.62	5.75	7	5
12F5V	6.45	2.89	12.52	7.23	9	8
13F1	.85	.35	4.50	4.75	7	10
14F4	.70	4.10	10.07	6.25	9	9
15M5	.85	.35	5.22	6.07	7	5
16F1V	5.27	4.40	10.70	7.27	8	7
17F4V	16.10	32.20	4.83	5.52	8	4
18F5V	13.90	5.35	8.70	7.97	6	7
19F1	20.00	3.10	8.37	7.15	9	7
20F4	3.40	18.40	3.87	8.30	8	6
21F1nV**	15.63	16.60	6.82	5.87	10	9
22F1nV	5.80	5.65	7.82	7.80	8	4
23F1nV	5.45	4.80	8.55	12.77	5	2

*Belt numbers 1, 4 and 5. Other numbers and letters have the same significance as in Table V-A.

**Belt 1nV, the first revision. rejected in favor of belts 2n and 3n.

APPENDIX C

THE GENERAL INFORMATION TEST

- | | |
|---|---|
| Y N ? 1. Is Valley Forge situated in New Jersey near Trenton? | Y N ? 27. Does the pressure of water depend upon its depth? |
| Y N ? 2. Did the Treaty of Ghent terminate the War of 1812? | Y N ? 28. Are the pampas found in the same regions as the tundras? |
| Y N ? 3. Is cotton gin an intoxicating alcoholic liquor? | Y N ? 29. Can tobacco be cultivated successfully in Northern states? |
| Y N ? 4. Did the Bolsheviks execute Kerensky after defeating him? | Y N ? 30. Does anyone question Shakespeare's authorship of "Hamlet"? |
| Y N ? 5. Does the "gourmand" differ essentially from the "gourmet"? | Y N ? 31. Are all mammals viviparous? |
| Y N ? 6. Do adjectives vary frequently in tense and mood? | Y N ? 32. Was President Garfield shot by the assassin Czolgosz? |
| Y N ? 7. Was Guinevere a character in "The Vision of Sir Launfal"? | Y N ? 33. Are the Californian redwoods the oldest living organisms? |
| Y N ? 8. Was "Gulliver's Travels" in any sense a political book? | Y N ? 34. Is the anteater an edentate animal? |
| Y N ? 9. Was Dickens interested in the condition of the poor? | Y N ? 35. Is Popocateptl in the same country as Chichen-Itza? |
| Y N ? 10. Does the allegory usually deal with symbolic characters? | Y N ? 36. Are there any nudists among the Dukhobors? |
| Y N ? 11. Was George Eliot's wife also a writer of fiction? | Y N ? 37. Was Paderewski ever noted as a statesman? |
| Y N ? 12. Is Winston Churchill the author of "Mr. Britling Sees It Through"? | Y N ? 38. Did Strauss compose the opera "Lohengrin"? |
| Y N ? 13. Is a straight line always the shortest distance between two points? | Y N ? 39. Is peonage a form of involuntary servitude? |
| Y N ? 14. Does the complement of an angle exceed its supplement? | Y N ? 40. Are terminal moraines of glacial origin? |
| Y N ? 15. If two planes intersect each other, is a line formed? | Y N ? 41. Was Vishnu one of the characters in Greek mythology? |
| Y N ? 16. If a triangle has two equal angles, is it isosceles? | Y N ? 42. Do all metals sink when placed on the surface of water? |
| Y N ? 17. Is the base of a triangle always the longest side? | Y N ? 43. Does Weber's Law apply to least noticeable differences? |
| Y N ? 18. Does combustion mean the same thing as explosion? | Y N ? 44. Are axones and dendrites identical in function? |
| Y N ? 19. Is the meter a longer unit than the yard? | Y N ? 45. Is asbestos cloth made from a mineral substance? |
| Y N ? 20. Did Kepler devise the earliest astronomical telescope? | Y N ? 46. Is the "displacement" of a boat related to its weight? |
| Y N ? 21. Does the hydraulic press operate by atmospheric pressure? | Y N ? 47. Does copper wire offer a high resistance to electricity? |
| Y N ? 22. Is Jupiter farther from the sun than the Earth? | Y N ? 48. Are spiders classified in the same phylum with crabs? |
| Y N ? 23. Is the ohm a unit of electrical capacity? | Y N ? 49. Did Longfellow write the poem entitled "Snowbound"? |
| Y N ? 24. Did the English take New York from the French? | Y N ? 50. Does the Government guarantee to redeem National Currency? |
| Y N ? 25. Do all birds make their nests in trees? | Y N ? 51. Did Samuel Johnson write an account of the life of Boswell? |
| Y N ? 26. Do glucose and sucrose belong to the same class of things? | Y N ? 52. Does all food depend ultimately upon photosynthesis? |
| | Y N ? 53. Is color-blindness generally believed to be hereditary? |

- | | |
|---|---|
| Y N ? 54. Is Wolfgang Köhler one of the better-known psychoanalysts? | Y N ? 82. Is Bogota the capital of the republic of Venezuela? |
| Y N ? 55. Do the Behaviorists deal extensively with subjective processes? | Y N ? 83. Was President Wilson's first name Thomas? |
| Y N ? 56. Is latitude measured perpendicular to the meridians? | Y N ? 84. Are rewards offered to those who report smugglers? |
| Y N ? 57. Does the spinal cord control involuntary reflexes? | Y N ? 85. Was Nathan Hale executed near the end of the Revolution? |
| Y N ? 58. Is an alloy a kind of musical instrument? | Y N ? 86. Did any part of the United States have woman suffrage fifty years ago? |
| Y N ? 59. Are proteins a necessary food element? | Y N ? 87. Did the great San Francisco fire and earthquake occur in 1910? |
| Y N ? 60. Is apathy synonymous with indifference? | Y N ? 88. Was "The Blue Boy" painted by Gainsborough? |
| Y N ? 61. Did General Pershing say, "Lafayette, we are here"? | Y N ? 89. Does "savoir faire" mean about the same as "nonchalance"? |
| Y N ? 62. Do Americans generally favor economic individualism? | Y N ? 90. Is a "de facto" government usually constitutional? |
| Y N ? 63. Are sensations more complex than perceptions? | Y N ? 91. Does a large thread-count indicate a high-quality fabric? |
| Y N ? 64. Does the hermit seek extensive social contacts? | Y N ? 92. Is Monte Carlo the capital of Monaco? |
| Y N ? 65. Does union labor favor the "closed shop" policy? | Y N ? 93. Did Columbus discover the island of Cuba in 1492? |
| Y N ? 66. Have all normal quadrupeds four locomotive appendages? | Y N ? 94. Is the mesencephalon a neurological structure? |
| Y N ? 67. Is the Liberian nation governed by a monarchy? | Y N ? 95. Was St. Paul one of the original twelve Disciples? |
| Y N ? 68. Are negroes practically disenfranchised in the South? | Y N ? 96. Are there any essential differences between a democracy and a republic? |
| Y N ? 69. Was Plutarch a famous Roman general? | Y N ? 97. Is the Mississippi the longest river in the world? |
| Y N ? 70. Was Rembrandt known as an early Renaissance scholar? | Y N ? 98. Does the term astigmatism refer to a religious belief? |
| Y N ? 71. Do marble and chalk have essentially the same chemical composition? | Y N ? 99. Is the astronomical light-year used to measure distance? |
| Y N ? 72. Are mutations believed to be caused by chemical changes in the genes? | Y N ? 100. When we multiply two numbers, is the result a quotient? |
| Y N ? 73. Do some believe that all matter is ultimately reducible to protons and electrons? | Y N ? 101. Is the word "supercede" spelled correctly? |
| Y N ? 74. Are proverbs and such sayings often mutually contradictory? | Y N ? 102. Is the radical always averse to change and progress? |
| Y N ? 75. Is Greenland commonly called the "Emerald Isle"? | Y N ? 103. Were Styx and Charon associated in Greek mythology? |
| Y N ? 76. Do sunspots always accompany terrestrial magnetic storms? | Y N ? 104. Does the U. S. Constitution make any reference to schools? |
| Y N ? 77. Can muscular strength be measured with a dynamometer? | Y N ? 105. Is the philatelist interested in things like postal rates? |
| Y N ? 78. Are platitudinous statements apt to be trite? | Y N ? 106. May an electrical battery have more than two poles? |
| Y N ? 79. Do some metals have low specific gravity? | Y N ? 107. Are bureaucratic governments rather likely to be encumbered with red-tape? |
| Y N ? 80. Can Congress reverse a Supreme Court Decision? | Y N ? 108. Do hallucinations have an actual sensory basis? |
| Y N ? 81. Is the St. Bernard Pass in the Pyrenees mountains? | Y N ? 109. Does USSR mean "United Soviet States of Russia"? |
| | Y N ? 110. Was "The Gold Bug" a tale of the California gold-rush? |

Y N ? 111. Does a chord join any two points on a circumference?	Y N ? 119. Is Gothic architecture characterized by the pointed arch?
Y N ? 112. Are postulates proved in geometrical reasoning?	Y N ? 120. Was Jean Valjean a character in "Les Miserables"?
Y N ? 113. Are transformers used to step up low-voltage currents?	Y N ? 121. Was Samuel Clemens' work noted for its humor?
Y N ? 114. Is the diameter of the earth the same in all directions?	Y N ? 122. Does a pastoral usually deal with country or farm life?
Y N ? 115. Was South Carolina the first state to secede from the Union?	Y N ? 123. Do two radii and an arc inclose a sector?
Y N ? 116. Does every State have fewer Senators than Representatives?	Y N ? 124. Is the angle of incidence always equal to the angle of reflection?
Y N ? 117. Was the Athenian Golden Age in the days of Pericles?	Y N ? 125. Is the referendum intended to safeguard the peoples' rights?
Y N ? 118. Did Montcalm command the French at the battle of Quebec?	

- - - - -

